université de BORDEAUX

THÈSE PRÉSENTÉE

POUR OBTENIR LE GRADE DE

DOCTEUR DE

L'UNIVERSITÉ DE BORDEAUX

ÉCOLE DOCTORALE : SOCIÉTÉS, POLITIQUES, SANTÉ PUBLIQUE

SPÉCIALITÉ : SANTÉ PUBLIQUE – OPTION EPIDÉMIOLOGIE

Par Timothy D'AOUST

Développement de stratégies de prévention de précision utilisant la prédiction du risque génomique pour la démence et les maladies cérébrovasculaires

Developing precision prevention strategies utilizing genomic risk prediction for dementia and cerebrovascular disease

Sous la direction de : Professeure Stéphanie DEBETTE

Soutenue le 16 décembre 2025

Membres du jury :

Ms. Stéphanie DEBETTE	Sorbonne University SIS (Sciences, Engineering, Health) (BORDEAUX)	Director of Thesis
Mr. Vincent MOOSER	Full professor, McGill University (Montreal, CANADA),	Rapporteur
Ms. Cornelia VAN DUIJN	Full professor, University of Oxford (UNITED KINGDOM)	Rapporteur
Ms. Hélène JACQMIN-GADDA	Professor, University of Bordeaux	President
Ms. Josée DUPUIS	Full professor, McGill University (Montreal, CANADA)	Examiner

Titre: Développement de stratégies de prévention de précision utilisant la prédiction du risque génomique pour la démence et les maladies cérébrovasculaires

Résumé: L'accident vasculaire cérébral (AVC) et la démence se classent respectivement au 3e et 8e rang des principales causes de décès en 2021, et la charge de ces affections devrait augmenter d'ici 2050, touchant principalement les pays à revenu faible et intermédiaire. Ces conditions sont liées à la **maladie cérébrale des petits vaisseaux (MPVC)**, caractérisée par des modifications des petits vaisseaux cérébraux, détectables par imagerie par résonance magnétique (IRM). Des études ont identifié environ 90 loci génétiques associés à l'AVC, plus de 75 à la démence, et plus de 50 liés aux marqueurs IRM de la MPVC. Les variants génétiques peuvent être combinés en scores de risque polygénique (PRS) qui quantifient la prédisposition génétique. Les PRS émergent comme des outils puissants pour guider la prévention et la gestion des maladies en stratifiant le risque et en prédisant le pronostic. Cependant, peu d'études ont validé ces outils dans des populations diverses, et leur valeur ajoutée par rapport aux modèles de risque clinique reste incertaine. Il existe une opportunité d'explorer de nouvelles approches de PRS qui capturent le risque polygénique associé à la pathologie vasculaire. Cette thèse visait à générer des preuves pour l'application des PRS dans les maladies cérébrovasculaires, en se concentrant sur l'AVC, la démence et la MPVC.

Nous avons utilisé des données cliniques et génétiques provenant de grandes cohortes et de l'essai clinique, PRESTIGE-AF. Dans la première partie, nous avons développé un PRS intégratif pour la démence (iPRS-DEM) intégrant des traits neurodégénératifs et vasculaires. L'iPRS-DEM a prédit la démence chez des personnes âgées de la communauté, des patients des cliniques mémoire, et dans une biobanque multi-ancestrale. Nous avons également montré que l'iPRS-DEM pouvait affiner la stratification du risque et être utile pour les essais cliniques et les programmes de prévention primaire.

Dans la deuxième partie, nous avons évalué les modèles de PRS dans l'essai PRESTIGE-AF chez des survivants d'hémorragie intracérébrale (HIC) avec fibrillation auriculaire (FA). Nous avons constaté que les PRS pouvaient fournir une stratification du risque cliniquement significative pour le traitement par anticoagulants oraux directs. Ces résultats soulignent que les PRS peuvent guider les stratégies de traitement dans des contextes d'incertitude.

Dans la dernière partie, nous avons appliqué des PRS aux hypersignaux de la substance blanche (HSB), le marqueur IRM le plus étudié de la MPVC. Nous avons montré qu'un PRS des HSB (HSB-PRS) était associé à l'AVC et à la démence. Grâce aux PRS basés sur les voies, nous avons identifié des voies biologiques potentiellement impliquées dans la MPVC. Nos

résultats ont mis en évidence des voies impliquées dans le métabolisme lipidique et la transduction du signal. Ce travail a démontré que les PRS pourraient améliorer la prédiction du risque de maladie et délimiter l'étiologie de la MPVC. Dans l'ensemble, cette thèse a généré des épreuves précieuses pour l'utilisation des PRS dans les maladies cérébrovasculaires et leur traduction en pratique clinique.

Mots clés: score de risque polygénique, prédiction du risque génétique, accident vasculaire cérébral, démence, maladie des petites artères cérébrales

Title: Developing precision prevention strategies utilizing genomic risk prediction for dementia and cerebrovascular disease

Abstract:

Stroke and dementia ranked 3rd and 8th, respectively, as leading causes of death in 2021, and the global burden of these conditions is projected to increase by two to threefold by 2050, mostly affecting low- and middle-income countries. Underlying both conditions is cerebral small vessel disease (cSVD), a common yet underrecognized pathology characterized by changes in the structure and function of small brain vessels detectable by brain magnetic resonance imaging (MRI). Genome-wide association studies (GWAS) have identified approximately 90 genetic loci associated with stroke, over 75 for dementia, and over 50 related to cSVD MRI markers. The effects of genetic variants from GWAS can be combined in polygenic risk scores (PRS) that quantify genetic predisposition to these conditions and associated traits. PRS are emerging as powerful genomic tools that have the potential to guide disease prevention and management by stratifying disease risk and predicting disease prognosis and response to therapy. In stroke and dementia, while PRS show predictive ability, key gaps remain: few studies have validated them longitudinally in diverse populations or clinical settings, and their added value beyond existing clinical risk models remains uncertain. There is also an opportunity to explore novel PRS approaches that more fully capture polygenic risk associated with vascular pathology or deconstruct genetic risk into biological pathways, particularly relevant for cSVD, for which mechanisms remain poorly understood. This thesis aimed to generate evidence for the application and refinement of PRS in cerebrovascular disease, with a focus on stroke, dementia, and cSVD.

We leveraged individual-level clinical and genetic data from large population-based cohorts and biobanks, a memory clinic-based cohort, and the PRESTIGE-AF clinical trial. In the first part, we aimed to improve dementia risk prediction and stratification by developing an integrative PRS for dementia (iPRS-DEM), incorporating both neurodegenerative and vascular traits. iPRS-DEM

predicted dementia across older community-persons, memory-clinic patients, and in a multi-ancestry biobank. We further demonstrated that alongside APOE ε4, iPRS-DEM may refine risk stratification and be particularly useful for the enrichment of clinical trials and primary prevention programs.

In the second part, we evaluated existing PRS models in the PRESTIGE-AF trial of intracerebral hemorrhage (ICH) survivors with atrial fibrillation (AF) who are at high risk of both thromboembolic and bleeding events. We found that PRS, particularly AF-PRS, could provide clinically meaningful risk stratification for direct oral anticoagulation treatment in these patients. These findings underscored that PRS has the potential to guide optimal treatment strategies in contexts of clinical equipoise where uncertainty is greatest. Such scenarios may represent the most pragmatic starting point for PRS implementation in secondary stroke prevention.

In the last part, we applied standard and pathway-based PRS to white matter hyperintensities (WMH), the most studied MRI marker of cSVD, in independent cohorts spanning the adult lifespan and clinical spectrum. We showed for the first time that a global WMH-PRS was associated with stroke and dementia, and through pathway-based PRS, we identified a set of biological pathways potentially implicated in cSVD across the lifespan. Our findings highlighted pathways involved in lipid metabolism, ciliogenesis, and signal transduction that may underlie hallmarks of cSVD pathology. This work demonstrated that PRS could be useful not only for improving disease risk prediction but also for delineating etiology in cSVD. Overall, this thesis generated valuable evidence for the use of PRS in cerebrovascular disease and their translation into clinical and public health practice.

Keywords : polygenic risk score, genetic risk prediction, stroke, dementia, cerebral small vessel disease

Unité de recherche

INSERM U1219 Bordeaux Population Health Center, Equipe ELEANOR
“Epidémiologie moléculaire des troubles vasculaires et cérébraux » Université de Bordeaux,
ISPED – 146 rue Léo Saignat – 33000 Bordeaux

Conditions optionnelles de réutilisation : *veuillez ne conserver que le logo de la licence appropriée*



Plus de détails dans le tutoriel ou en suivant ce lien https://creativecommons.org/licenses/?lang=fr-FR

Acknowledgements

To **Stéphanie Debette**, my thesis supervisor, thank you for these past four years. I am incredibly grateful to work alongside you on such innovative projects, and for the opportunity to participate in international collaborations, and attend and present at such prestigious conferences. The opportunities you have provided will leave a lasting impact and has fostered an interest in research that I didn't know I possessed. I admire your scientific vision, as well as your inexhaustible work ethic and commitment. Thank you for your guidance during this PhD journey.

To the **members of the jury.** I thank **Rodolphe Thiébault** for accepting the position as president of my jury. I wouldn't be here without your efforts to build the public health data science program, and I am honored that you would oversee my defense as the culmination of my work that started during this master's program. I also extend my gratitude to the rapporteurs of my thesis, **Cornelia van Duijn** and **Vincent Mooser**, I was honored to have you accept the invitation to read and critique my thesis, I hope you found it worth your time. Additional thanks to **Josée Dupuis** first for your encouragement, understanding, and valuable advice during my committee meetings, and now for your willingness to be part of my jury. Thank you to **Hélène Jacqmin-Gadda** for agreeing to be on my jury and for your help particularly to improve my biostatistical knowledge.

I also thank the **Digital Public Health program** for financing this thesis, and for cultivating an enriching learning environment within the program. I also extend my thanks to **RHU SHIVA** (ANR-18-RHUS-0002) and the **l'IHU VBHI** (ANR-23-IAHU-0001) for fostering a dynamic research environment for the duration of my thesis.

Aux members de l'equipe ELEANOR, je vous remercie beaucoup de m'avoir encadré pour les dernières quatres années. Je vous remercie encore plus d'avoir accueilli un Canadien anglophone qui ne parlait aucun français au début… Je suis fier que, désormais, vous ne passiez plus à l'anglais quand j'arrive en réunion. Merci également pour tous les échanges scientifiques et les échanges conviviaux.

> À **David**, merci pour toute l'attention et tous les efforts que tu fournis chaque jour pour garantir un environnement accueillant et dynamique au sein de l'équipe. En tant qu'étranger, cela fait toute la différence. Merci pour ton soutien pendant mon CSI, ainsi que pour ta gentillesse et ta disponibilité constantes. J'admire beaucoup ton leadership et ton esprit scientifique, et je suis ravi de continuer à échanger avec toi.

À **Cécilia**, merci pour ta gentillesse et pour toujours avoir été très chaleureuse avec moi. Je te serai toujours reconnaissant pour ta compréhension et ta disponibilité. J'admire la curiosité et l'humilité avec lesquelles tu abordes la science — des qualités auxquelles j'aspire moi-même. Merci de m'avoir fait confiance et de m'avoir accueilli comme ton postdoc; j'ai hâte de travailler avec toi pour la suite.

À **Morgane**, merci pour ton inépuisable gentillesse, pour ta résilience, et bien sûr pour ton aide avec PRESTIGE-AF et toutes les autres questions logistiques. Je sais que je peux compter sur toi.

À **Judith**, merci pour ton énergie toujours chaleureuse — tu m'as beaucoup manqué au bureau. Je suis très heureux que tu sois revenue.

À **Lucile** et **Alice**, merci pour votre patience face à toutes mes difficultés logistiques. Vous êtes bien plus organisées que je ne pourrai jamais l'être.

To **Aniket**, thank you for all the scientific discussions, I wish you all the best in Paris.

À l'ensemble des jeunes chercheur·e·s de l'équipe (actuel ou passé): **Ilana, Coralie, Blandine, Ima, Floriane, Constance, Emmanuelle, Audrey, Mathilde, Jeanne, Florian, Quentin, Gaëlle, Sophie, Hema, Akshara et Amélie** — j'espère n'avoir oublié personne. Merci d'avoir toujours été accueillant·e·s et pour votre patience avec mon français. Merci pour toutes les discussions et pour les petits échanges qui illuminent les journées. A particular thanks to **Quentin** for always having the time for honest scientific discussion, your presence on the team was always valuable and you'll find great success in research, I'm sure. Un merci particulier également à **Blandine**, qui m'a toujours témoigné une gentillesse exceptionnelle, surtout quand j'en avais le plus besoin. À **Constance**, merci pour ton soutien inépuisable, et surtout pour tes incroyables gâteaux au chocolat. And finally, **Ilana,** you've been by my side from the beginning. Thank you for being an amazing colleague and an even better friend. You've been a huge part in helping me through this journey, and I know I can always count on you. Your potential is endless, just like your positive energy.

To international collaborators, in PRESTIGE-AF and the CHARGE neurology working group, it was a pleasure to meet in person and to work alongside you, thank you for your advice and support.

To **Rebecca Fuhrer**, thank you for all your support and for really facilitating this experience in Bordeaux through the public health data science program. You deeply care

for the students in this program, and that's what makes it such a success. You're always there to support my concerns and aspirations.

To my friends in Bordeaux:

À tout le monde au CAM, vous ne le saviez peut-être pas, mais vous m'avez sauvé au cours de ces dernières années. Le CAM est ma place, et je ne peux pas décrire avec des mots (surtout en français) à quel point cela a été précieux pour moi.

To **Julien**, you're like the big brother I never had! It changed everything when I decided to start destroying myself on Sunday Hyrox with you. We share many interests, values, and laughs, and even almost the same birthday. I'm very grateful for our friendship.

To **Mat**, you're the other big brother! You've been that constant guy from the beginning, always there for me through whatever. Thanks for always making me cry of laughter, and always doing you.

To **Arash**, thanks for being the guy I can always count on for a cold beer and some great chats about the gym, being a foreigner in France, the challenges of academia, and life.

À **Marie-Édith**, je ne pourrai jamais te remercier assez pour tout ce que tu as fait pour moi depuis le début de mon séjour à Bordeaux. Merci de m'avoir hébergé pendant mon master et le début de ma thèse, et pour tous les dimanches midi que nous avons passés ensemble. Tu as toujours été une constante, toujours là pour moi, et tu es devenue une partie intégrante de ma vie ici. Je te considère comme un membre de ma famille. Merci pour tout.

To my friends outside of Bordeaux:

To **Spencer**, you're my dude. We've experienced so much and grown together, and I'm grateful that even though I've been far it feels like we always find a way to reunite. You're truly a best friend. I don't think I could have done this without those three months in Montreal with you and **Erica**. Erica, you're an absolute sunshine and I'm so grateful for the friendship we have both as a trio with Spencer and between us. A huge thanks to your family as well, the **Carter-Williamson's**, for welcoming for Christmas 2024 as part of their family. I was so honored to share in the love that you share with each other. Spencer and Erica, I love you guys! Big hugs!

To **Austin,** I am very grateful for rekindling our high school friendship over the past years. Our long chats on the phone were always a treat. I appreciate never knowing where our

conversations can lead, but they often go to amazing places. You've offered a lot of support for me even from afar, keep doing your thing.

To **Benji,** from the moment we met on a racous night in Laos we knew there was a connection. I've been extremely grateful to be closer to you in Europe over the past few years, and to share some quality time together. From the France van trip trying to find good coffee (we didn't) to the great coffees we had in Italy together, and spending the holidays together and being welcomed into your family. Crib games, coffee, and never-ending laughter and love! There's not more I could ask for. I look forward to every minute of our next phone call and especially when we can be together again. Your friendship has been a rock for me, thank you for everything.

To **Emma**, mon stein. There are so many things I could list that I love and appreciate about you and our relationship, but I think I can sum it up by saying, we are stein, you deeply get me and I deeply get you and that's the most important thing. Your support of me during this thesis goes beyond words and beyond cookies and chocolate. I could get used to this stein life. Je t'aime beacuoup.

To my family:

To my **Grandad and Grandmaman** thank you for your perpetual love and kindness. I will always treasure the time I've got to spend in Montreal with you over the past years. Thank you for always supporting me, I love you both.

To **Grandma Paddy and Grandpa Nicky,** I felt your support and encouragement from Mexico, and I still feel it from Victoria. Soon you can tell your friends that you have a "Doctor" in the family. I love you both.

To **Emerson**, my little brother, we've been apart but still our relationship has grown over the past few years. I'm grateful for that and it's great to see you now doing your own thing. Thank you for holding down the fort with Mum. I love you, bro!

To **Mum**, your strength has supported me since day one. It's difficult being so far, but it makes the time we spend together so much more precious. Your visits to France are treasured, and it means so much that you make one more to be here for my thesis defense. Your love and support always keeps me going. I love you, mum!

To **Dad**, our relationship is a pillar in my life. You've been my number one fan through this all, and the cosmos wouldn't have it any other way! I love you, Dad!

Table of Contents

Scientific Production 10
List of Tables and Figures 13
Abbreviations 14
General Introduction 16
Chapter 1 – Cerebrovascular Diseases 18
1.1 Stroke and dementia in the context of an aging world 18
1.2 Brain Structure and Vasculature 20
1.2.1 Basic Brain Anatomy 20
1.2.2 Blood supply of the brain 22
1.2.3 The Neurovascular Unit 24
1.3 The spectrum of cerebrovascular disease 25
1.3.1 Cerebral Small Vessel Disease 26
1.3.2 Stroke 30
1.3.3 Dementia 33
Chapter 2 – Genetic Risk Prediction 41
2.1 Introduction to Genetic Risk Prediction 42
2.1.1 A brief history of genetic epidemiology 42
2.1.2 GWAS and heritability 45
2.1.3 Polygenic risk scores 48
2.2 Genetic Architecture of Cerebrovascular Diseases 50
2.2.1 Monogenic disorders 50
2.2.2 Common genetic variation in cSVD 50
2.2.3 Common genetic variation in stroke 51
2.2.4 Common variation in dementia 52
2.2.5 A note on SNP-based heritability estimations 53
2.3 Polygenic Risk Scores Applied to Cerebrovascular Disease 54
2.3.1 PRS for cSVD MRI-markers 55
2.3.2 Stroke PRS 55
2.3.3 Dementia PRS 56
2.3.4 Potential utility of PRS 58
2.3.5 Caveats and challenges to PRS implementation 60
Rationale and Aims of Thesis 61
Chapter 3 – Methods 64
3.1 Study Populations 65
3.1.1 Three City Study (3C) 65
3.1.2 MEMENTO 67
3.1.3 i-Share 69
3.1.4 PRESTIGE-AF 71
3.1.5 UK Biobank 72
3.1.6 All of Us Research Program 74
3.2 Survival analysis and competing risk models 75
3.2.1 Choice of time scale 75
3.2.2 Censoring 76
3.2.3 Competing Risks 77
3.2.4 Cause-specific hazard vs. Fine-Gray Regression models 78
3.3 Polygenic Risk Scores 79
3.3.1 Principals of constructing PRS 79
3.3.2 Standard PRS Methods 81
3.3.3 Advanced PRS methods 82
3.3.4 Optimizing PRS 88
3.3.5 Investigating PRS Associations 90

Chapter 4 – Integrative PRS for dementia ____ *95*
Chapter 5 – PRS to inform secondary stroke prevention ____ *115*
Chapter 6 - Pathway-based PRS approach to cSVD ____ *144*
Chapter 7 – Discussion ____ *180*
7.1 Summary of Findings ____ **180**
7.2 Interpretation of Findings ____ **182**
7.2.1 An integrative approach shows ability to stratify dementia risk ... 182
7.2.2 PRS for cerebrovascular disease likely provide greater utility in high-risk populations. ... 185
7.2.3 The beginning of translatable PRS for cSVD ... 189
7.3 Limitations and considerations to improve PRS risk prediction ____ **191**
7.3.1 Generalizability across non-European ancestries ... 192
7.3.2 Lack of standardization in PRS calculation and reporting ... 193
7.3.3 Limitations in study design ... 194
7.3.4 Criticism of the polygenic additive model ... 196
7.3.5 Barriers to clinical implementation ... 198
7.4 Perspectives and Future Directions ____ **200**
7.4.1 Future Directions: Epistatic effects, lifestyle and environment ... 200
7.4.2 Perspective: Effective risk communication is essential to PRS utility ... 202
7.5 General conclusion ____ **204**
References ____ *205*
Supplementary Appendices ____ *225*

Scientific Production :

Publications

1.2 Published

D'Aoust, T. et al. Polygenic score integrating neurodegenerative and vascular risk informs dementia risk stratification. **Alzheimer's & Dementia** 21, e70014 (2025).
Contribution of the PhD candidate: All analyses apart from the external validation in All of Us in which case a detailed analysis plan was provided to the analyst, interpretation of the results writing of the full manuscript

Arnould L*, Soumaré A*, **D'Aoust T***, et al. Retinal Microvascular Correlates of Cerebral Small Vessel Disease in Older Age. **Neurology**. 2025;105(9):e214251. doi:10.1212/WNL.0000000000214251* All authors contributed equally to this work
Contribution of the PhD candidate: mendelian randomization analysis, writing of part of the manuscript (results, methods, discussion), revision of the manuscript.

Caro I, Western D, Namba S, Sun N, Kawaguchi S, He Y, Fujita M, Roshchupkin G, **D'Aoust T**, Duperron MG, Sargurupremraj M, Tsuchida A, Koido M, Ahmadi M, Yang C, Timsina J, Ibanez L, Matsuda K, Suzuki Y, Oda Y, Kanai A, Jandaghi P, Munter M, Auld D, Astafeva I, Puerta R, Rotter JI, Psaty BM, Bis JC, Longstreth WT, Couffinhal T, García-González P, Pytel V, Marquié M, Cano A, Boada M, Joliot M, Lathrop M, Le Grand Q, Launer LJ, Wardlaw JM, Heiman M, Ruiz A, Matthews PM, Seshadri S, Fornage M, Adams H, Mishra A, Trégouët DA, Okada Y, Kellis M, De Jager PL, Tzourio C, Kamatani Y†, Matsuda F†, Cruchaga C†, Debette S†. Proteogenomics in cerebrospinal fluid and plasma reveals new biological fingerprint of cerebral small vessel disease. Submitted. Preprint, doi: 10.21203/rs.3.rs-4535534/v1. PMID: 39011113. [**accepted in Nature Aging**].
Contribution of the PhD candidate: PRS analysis, writing of methods and results of respective section, revision of the manuscript.

1. Veltkamp R, Korompoki E, Harvey KH, Harvey ER, Fießler C, Malzahn U… **D'Aoust T**… et al. Direct oral anticoagulants versus no anticoagulation for the prevention of stroke in survivors of intracerebral haemorrhage with atrial fibrillation (PRESTIGE-AF): a multicentre, open-label, randomised, phase 3 trial. **The Lancet.** 2025 Mar 15;405(10482):927–36.
Contribution of the PhD candidate: Participation in consortium annual meetings, quality control and analysis of data as part of genetic sub-study, revision of the manuscript.

Harvey, K. H. ; K., Eleni; Harvey, Emily R. ;. Fießler, Cornelia; Malzahn, Uwe; Hügen, Klemens; Ullmann, Sabine; Schuhmann, Carolin; Putz Todd, Gabriele; Montaner, Joan; Penalba, Anna; Guaman-Pilco, Daisy; Sibon, Igor; Debette, Stephanie; **D'Aoust, Timothy**; Lachaize, Morgane; Enzinger, Christian; Ropele, Stefan; Fandler-Höfler, Simon; Ruecker, Viktoria; Haas, Kirsten; Nielsen, Peter B. ;. Wolfe, Charles; Wang, Yanzhong; Wafa, Hatem; Caso, Valeria; Mosconi, Maria Giulia; Lip, Gregory Y. H. ;. Lane, Deirdre A. ;. Haefeli, Walter E. ;. Foerster, Kathrin I. ;.

Wurmbach, Viktoria S. ;. Ringleb, Peter; Heuschmann, Peter U. ;. Veltkamp, Roland; on behalf of PRESTIGE-AF investigators. Personalized Medicine, Public Health and Patient-Centred Aspects in the Prevention of Stroke in Intracerebral Haemorrhage Survivors with Atrial Fibrillation (PRESTIGE-AF) Project. **Thromb Haemost** https://doi.org/10.1055/a-2576-7760 (2025) doi:10.1055/a-2576-7760.
Contribution of the PhD candidate: Participation in consortium annual meetings, quality control and analysis of data as part of genetic sub-study, revision of the manuscript.

1.2 Submitted or under revision

Osorio Meirelles[1,#], Claudia P Cabrera[2,3,#], Zhiguang Li[1], Valborg Gudmundsdottir[4,5], **Timothy D'Aoust[6]**, , Jie Yao[xx], Xiuqing Guo[xx], Timothy Hughes[xx], Nicole Benfield[2], Ajay K Gupta[2,9], Morris Brown[2], Mike A Nalls[7,8], Stephanie Debette[6], Vilmundur Gudnason[4,5], Michael R Barnes[2,3], Patricia B Munroe[2,3], Lenore J Launer[1#*], Helen R Warren[2,3#*]. Pathway-based genetic risk score analyses identify biological pathways linking blood pressure and brain-MRI traits. [Submitted to American Journal of Human Genetics].
Contribution of the PhD candidate: analysis of 3C-Dijon and i-Share data, revision of manuscript.

1.3 In preparation

Tim D'Aoust*[5], Shufan Huo, Santiago Clocchiatti-Tuozzo, Peter U. Heuschmann[1,2], Joan Montaner[3,4], Jemma Hopewell, Morgane Lachaize[5,6], Guillaume Paré[7], Charles Wolfe[8, 9], Guido Falcone, Roland Veltkamp[10, 11, 12], Stéphanie Debette[5, 6] . Polygenic risk informs anticoagulation treatment in intracerebral hemorrhage survivors with atrial fibrillation. [in preparation for *Circulation*]
Contribution of the PhD Candidate: all statistical analyses, interpretation of results, and writing of full manuscript.

Tim D'Aoust (MSc, Université de Bordeaux) ; Carole Dufouil (PhD, Université de Bordeaux) ; Ami Tsuchida (PhD, Université de Bordeaux) ; Quentin Le Grand* (PhD, Université de Bordeaux) ; Stéphanie Debette* (PhD, Université de Bordeaux). Pathway-specific polygenic risk scores reveal underlying mechanisms of cerebral small vessel disease. [in preparation for submission to *Brain*]
Contribution of the PhD Candidate: all statistical analyses, interpretation of results, and writing of full manuscript.

Oral and poster presentation

2.1 Oral Presentations

European Stroke Organization Conference (ESOC). May 2025. Helsinki, Finland. Polygenic risk scores have potential to stratify stroke risk and inform direct anticoagulant use in ICH survivors with atrial fibrillation.

Cohorts of Hearts and Aging Research in Genomic Epidemiology (CHARGE) consortium meeting. May 2025. Poster blitz participant. *Washington D.C., USA.* Pathway-based polygenic risk scores reveal mechanisms of cerebral small vessel disease across the lifespan

European Stroke Organization Conference (ESOC). May 2024. *Basel, Switzerland***.** An integrative polygenic risk score for all-cause dementia shows potential for risk stratification across population and clinical settings.

International Stroke Genetics Consortium Workshop. May 2024. *Munich, Germany***.** An integrative polygenic risk score for all-cause dementia shows potential for risk stratification across population and clinical settings.

2.3 Poster Presentations

European Stroke Organization Conference (ESOC). May 2025. *Helsinki, Finland***.** Pathway-based polygenic risk scores reveal mechanisms of cerebral small vessel disease across the lifespan.

Cohorts of Hearts and Aging Research in Genomic Epidemiology (CHARGE) consortium meeting. May 2023. Poster blitz participant. *Boston, USA***.** An integrative polygenic risk score for all-cause dementia shows potential for risk stratification across population and clinical settings.

List of Tables

Table 1. *Description of MRI markers of cSVD.* ***27***
Table 2. *Overview of dementia aetiologies.* ***35***
Table 3. *APOE global prevalence* ***52***
Table 4. *Proposed clinical utility of PRS in cardiometabolic disease* ***59***
Table 5. *Summary of PRS limitations* ***192***

List of Figures

Figure 1. *Compression of morbidity* ***19***
Figure 2. *Representation of a neuron and a coronal section of the brain* ***21***
Figure 3. *Anatomical brain organization* ***22***
Figure 4. *Brain arterial system* ***23***
Figure 5. *Brain venous system* ***24***
Figure 6. *The Neurovascular Unit* ***25***
Figure 7. *Iceberg model to describe the spectrum of cerebrovascular disease* ***26***
Figure 8. *MRI markers of covert cSVD* ***27***
Figure 9. *Conceptual relationship between AD, VCI and VaD* ***37***
Figure 10. *Lancet Commission modifiable risk factors of dementia* ***38***
Figure 11. *Example of a Manhattan plot from GWAS* ***46***
Figure 13. *Potential of PRS to be used for population stratification* ***58***
Figure 14. *Overview of study populations* ***65***
Figure 15. *Polygenic risk score distribution* ***80***
Figure 16. *The 'pathway polygenic' risk score approach* ***85***
Figure 17. *PRS case-control plots* ***91***
Figure 18. *Two common ways to categorize PRS* ***92***
Figure 19. *Factorial PRS design to stratify into clinically-relevant groups* ***94***
Figure 20. *10-year cardiovascular risk 2021 ESC guidelines* ***183***

Abbreviations

h^2_{SNP}: SNP-based heritability
3C: Three Cities Study
ACD: All-cause dementia
AD: Alzheimer's Disease
ADNI: Alzheimer's Disease Neuroimaging Initiative
AF: Atrial Fibrillation
AoU: All of Us Research Program
APOE: apolipoprotein E
APP: amyloid precursor protein
ARIA: Amyloid-related imaging abnormalities
ARR: absolute risk reduction
AUC: area under the curve
Aβ: amyloid-beta or beta-amyloid
BBB: blood-brain-barrier
BMI: body mass index
BHS: brain health services
C+T: clumping and thresholding
CAA: cerebral amyloid angiopathy
CAD: coronary artery disease
CADASIL: cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy
CARASIL: cerebral autosomal recessive arteriopathy with subcortical infarcts and leukoencephalopathy
CES: cardioembolic stroke
CIF: cumulative incidence function
CMB: cerebral microbleeds
CNS: central nervous system
CSF: cerebral spinal fluid
CSH: cause-specific hazard model
cSVD: cerebral small vessel disease
dMRI: diffusion-based MRI
DOAC: direct oral anticoagulants
DSM-(IV/V): Diagnostic and Statistical Manual of Mental Disorders, Fourth or Fifth Edition
DTC-GT: Direct-to-consumer genetic testing
DTI: diffusion-tensor imaging
EADB: European Alzheimer's & Dementia Biobank
ECM: extracellular matrix
EHR: electronic health records
ESC: European Society of Cardiology
ESO: European Stroke Organization
FDA: Food & Drug Administration (USA)
FGR: Fine and Gray Regression
FTD: frontotemporal dementia
GRS: genomic risk score
GWAS: genome-wide association study
GWS: genome-wide significance
GxE: gene by environment interaction
HR: hazard ratio
HRC: Haplotype Reference Consortium
IBD: identity by descent
ICD: International classification of disease
ICH: intracerebral hemorrhage
INFO: imputation quality
iPRS: integrative polygenic risk score
IS: ischemic stroke
IS-iPRS: ischemic stroke integrative polygenic risk score
LAS: large artery stroke
LBD: Lewy body dementia
LD: linkage disequilibrium
LDL: low-density lipoprotein
LDSC: linkage disequilibrium score regression
LMIC: low- and middle-income country
MACE: major adverse cardiovascular event
MAF: minor allele frequency
MCI: mild cognitive impairment
MR: Mendelian randomization
MRI: magnetic resonance imaging
NFT: neurofibrillary tangles
NGS: next-generation sequencing
NHS: National Health Service (UK)

NODDI: neurite orientation dispersion and density imaging

NRI: net reclassification index

NVU: neurovascular unit

OCPS: Office of Population Censuses and Surveys Classification of Interventions and Procedures

OR: odds ratio

PAL: polygenic additive liability

PC: genetic principal component(s)

PDD: Parkinson's disease dementia

PET: positron emission technology

PRS: polygenic risk score

pQTL: protein quantitative trait loci

PSMD: peak skeletonized mean diffusivity

ps-PRS: pathway-specific or pathway-based polygenic risk score(s)

P-tau: phosphorylated tau

PVS: perivascular spaces

RCT: randomized control trial

SAH: subarachnoid hemorrhage

SBP: systolic blood pressure

SD: standard deviation

SDI: socio-demographic index

sHR: sub-distribution hazard ratio

SNP: single nucleotide polymorphism

STRIVE: standards for reporting vascular changes on neuroimaging

SVS: small vessel stroke

T2D: type-2 diabetes

UKBB: UK Biobank

VaD: vascular dementia

VCI: vascular cognitive impairment

VMD: vascular or mixed dementia

WES: whole exome sequencing

WGS: whole genome sequencing

WHO: World Health Organization

WMH: white matter hyperintensities

General Introduction

In 2025, people are living longer than ever before, leading to a global shift towards non-communicable diseases—particularly age-related conditions such as stroke and dementia. Among non-communicable diseases, stroke and dementia represented the 2nd and 4th most common causes of death in 2021, with stroke affecting an estimated 94 million and dementia 57 million people worldwide (1,2). By 2050, we are projected to have over 150 million cases of both stroke and dementia, disproportionately arising from low-and-middle income countries (LMICs) (2,3). This rise is driven by global population aging and also increasing prevalence of vascular risk factors—including hypertension, obesity, and diabetes—that strongly influence the risk of both conditions.

These neurological conditions are complex and deeply interconnected. It is increasingly understood that the majority of dementia cases have a "mixed" pathology, involving both neurodegenerative and vascular brain lesions, including those caused by stroke (4–6). Further, a key underlying pathology is cerebral small vessel disease (cSVD), a common yet underrecognized disease characterized by changes in the structure and function of small brain vessels. cSVD is a major cause of stroke and the principal vascular substrate underlying cognitive decline and dementia (7,8). As such, targeted prevention of cSVD and the management of shared modifiable risk factors—particularly hypertension—offer major opportunities to prevent stroke and dementia in the general population (9).

In the era of precision medicine and public health, there is a growing recognition of the complex interplay between genetics, lifestyle, and the environment in shaping disease risk. Advances in cost-effective genomic technologies have enabled large-scale profiling of the germline genome, generating rich datasets that can be used to inform precision prevention strategies. Genome-wide association studies (GWAS) have leveraged these datasets to identify ~90 genetic loci associated to stroke (10,11), over 75 loci with dementia (12), and more than 70 loci linked to cSVD (13–16). Using data generated from GWAS, we can then quantify individual genetic predisposition to these conditions and associated traits through aggregating the effects of many common genetic variants in polygenic risk scores (PRS). PRS are emerging as promising genomic tools that have potential to guide disease prevention and management by stratifying disease risk and predicting disease prognosis and response to therapy (17,18).

In the context of cerebrovascular disease, PRS has shown ability to predict both stroke and dementia (19–22). However, key gaps remain: few studies have validated these tools longitudinally in diverse populations or real-world clinical settings, and their added value beyond

existing clinical risk models remains uncertain. There is also an opportunity to explore novel PRS approaches that more fully capture polygenic risk associated with vascular pathology. This thesis aims to generate critical evidence for the application and refinement of PRS in cerebrovascular disease, with a focus on stroke, dementia, and cSVD. Ultimately, the findings of this work will support the timely translation of genomic tools into clinical and public health practice and further advance the development of precision prevention strategies for age-related brain diseases.

Chapter 1 – Cerebrovascular Diseases

1.1 Stroke and dementia in the context of an aging world

The prevention of stroke and dementia is central to ensuring healthy longevity and quality of life among older people. Due to population aging, there is an ongoing global shift from infectious to chronic, age-related diseases such as stroke and dementia (3,23,24). Indeed, one in three healthy persons over the age of 55 will develop stroke or dementia during their remaining lifetime (25). Demographic and developmental trajectories, particularly in LMICs, assure the global burden of stroke and dementia will remain high.

In 2019, there were 1 billion people aged 60 years and older worldwide, a figure projected to more than double to 2.1 billion by 2050, increasing from 12% to 22% of the global population (26). Population aging occurs because as countries develop, urbanization, better education, women empowerment, and access to family planning among other factors lead to a decline in family size and fewer births (27). In parallel, life expectancy has risen largely through reductions in infant and child mortality, supported by improvements in healthcare, sanitation, nutrition, food security, and infectious disease control (e.g., vaccination programs) (27). Historically, since the 1950s, we have lived through a "longevity revolution" with global life expectancy rising from 47 to 73 years of age (28). While this represents a major public health achievement, most of these gains have resulted from reductions in early-life mortality, and additional increases in longevity are likely to continue to yield diminishing returns (28). Longer life naturally exposes more people to an elevated risk of aging-related diseases and more years lived in disability (28).

Reductions in mortality alone without proportional reductions in morbidity inevitably increase the burden of chronic disease at older ages. In response to population aging, it has been argued that modern medicine and public health should prioritize *healthspan*—the number of years lived free of disability—over *lifespan*, as the main measure of success (28). As early as 1980, James Fries proposed the concept of "compression of morbidity", stating that our goal should be to delay or compress the period of lifespan when frailty and disability increase substantially (29) (**Figure 1**). This is accomplished by preventing or delaying the onset and progression of age-related chronic

diseases, otherwise, reductions in mortality—particularly among older adults— may prolong unhealthy lifespan and worsen overall population health.

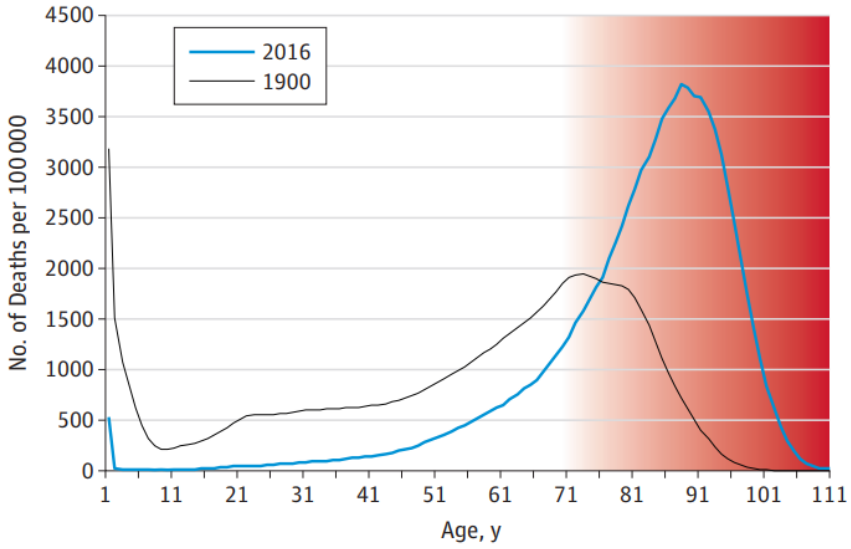


Figure 1. The age distribution of life table deaths for women in the United States per 100,000 people in 1900 versus 2016. The red zone represents the period of life when frailty and disability begin to rapidly increase. Compression of morbidity relates to delaying and compressing this red zone. *Figure adapted from Olshansky (2018)* (28)

Overall, most evidence indicates that morbidity has not significantly compressed to date (30) and increases in lifespan have not been accompanied by proportional gains in healthspan (30,31). Globally, life expectancy has risen faster than health-adjusted life expectancy, creating a current gap of 9.6 years that has continued to widen in recent decades (31). These lagging healthspan trends, combined with population aging, underpin projections that more than 300 million people will be affected by either stroke or dementia by 2050 (3,23).

Prevention of age-related conditions, such as stroke and dementia, is central to promoting healthy aging, and maintaining sustainable healthcare systems. Age-specific declines in prevalence of heart disease and stroke, alongside reductions in rates of mild disability and cognitive impairment, suggest improvements in the innate health of older people (30,32). In most developed countries, these promising trends are often attributed to better physical development early in life, societal improvements like better education and reduced physical demands of labor, less smoking, and improved management of key physiological indicators such as blood pressure through increased medication use (30,33). At the same time, however, there is a growing burden worldwide of metabolic diseases like obesity and diabetes (34,35), poor control of hypertension in LMICs (36), and limited evidence of improvements in severe physical disability (30).

The development of age-related chronic diseases can be viewed as a gradual, lifelong trajectory in which genetic, environmental, and lifestyle factors accumulate over time, driving physiological dysregulation, subclinical pathology, and ultimately overt disease, frailty, disability, and death

(37). Effective prevention must encompass both delaying the onset of disease (primary prevention) and slowing progression and disability among those already affected (secondary prevention). Prevention strategies should proactively target modifiable risk factors and incorporate early assessments of disease risk, ideally before the onset of morbidity. Concurrently, a deeper understanding of disease mechanisms can identify actionable targets, enhancing efforts to delay onset and slow progression. Stroke and dementia share several early and mid-life risk factors, many of which are modifiable; however, our ability to identify high-risk individuals and our understanding of disease etiology remain limited. The following sections will describe cerebrovascular diseases, beginning with an overview of brain structure and vasculature. We will then discuss their pathology, public health context, and current prevention strategies, before transitioning to genetic risk prediction in the next chapter.

1.2 Brain Structure and Vasculature

1.2.1 Basic Brain Anatomy

The brain is one of the most extraordinary and complex organs of the human body, serving as our main interface with the world and allowing us to perceive, think, feel emotions, communicate, and carry out actions. The brain is highly organized and structured into distinct and interconnected areas, comprising an intricate network of approximately 85 billion neurons that control essentially all bodily functions, along with an equivalent number of glial cells responsible for brain maintenance and homeostasis (38). The entire field of neuroscience is dedicated to studying the complexities of the brain and nervous system—from molecular and cellular mechanisms to cognition, behavior, and consciousness. This also involves extensive work to visualize brain structure and activity through advanced neuroimaging techniques. In this section I provide a brief overview of brain anatomical and functional organization, with a specific focus on brain vasculature and the neurovascular unit.

The brain is fundamentally divided into grey and white matter, each representing distinct features of neurons (**Figure 2**). Grey matter primarily consists of neuron cell bodies, dendrites, and synapses, which are largely dedicated to information processing (39). In contrast, white matter is composed of axons, often myelinated, that group together in tracts to rapidly transfer signals across different brain areas. Myelin, a principal component of white matter, accounts for about 50-60% of its weight (40). It is highly conductive, increasing conduction velocity by approximately 50 times compared to unmyelinated axons, thereby facilitating the structural connectivity essential for human cognition and behavior (40,41).

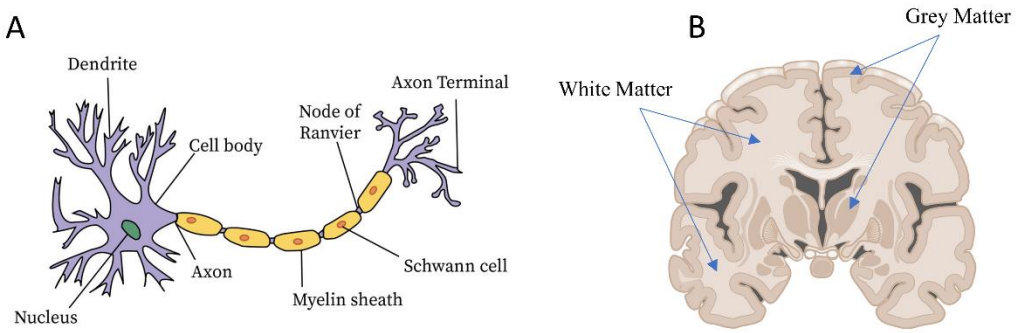


Figure 2. (A) Representation of a neuron and its main components. Grey matter is comprised of neuron dendrite, cell body, and nucleus on one end and the axon terminal on the other, forming synapses with other neurons. White matter comprises the axon which is surrounded by a myelin sheath and resident Schwann cell-- glial cells that produce and maintain the myelin sheath. (B) Coronal section of the brain showing the division of the brain into white and grey matter of the deep brain nuclei and cerebral cortex. Neuron image: CC BY-SA 3.0 | Image credits : https://commons.wikimedia.org | Anatomy and Physiology" by the US National Cancer Institute's Surveillance, Epidemiology and End Results (SEER). Coronal section of the brain: Slagter - Drawing Coronal section of the brain - no labels" at AnatomyTOOL.org by Ron Slagter, license: Creative Commons Attribution-NonCommercial-ShareAlike,

The anatomical structure of the human brain is composed of three primary divisions: the large cerebrum, the cerebellum, and the brain stem (42). The cerebrum, the largest part, consists of paired cerebral hemispheres whose surfaces are highly convoluted, featuring gyri (bumps) and sulci (grooves) that make up the cerebral cortex. These hemispheres are connected by the corpus callosum, a huge bundle of white matter tracts. White matter tracts extending from the cortex connect it to subcortical structures in the brain stem and spinal cord. The cerebrum is conventionally divided into four lobes: the frontal lobe, parietal lobe, the temporal lobe, and the occipital lobe, situated at the posterior of the cerebrum (**Figure 3A**). A buried piece of the cerebral cortex, known as the insula, separates the temporal and frontal lobes.

In the cerebrum, gray matter forms a thin outer sheet of the cerebral cortex which surrounds the white matter lying beneath. Various areas of the cerebral cortex in each of the lobes is dedicated to specific functions, such as voluntary movement (frontal lobe), and decoding sensory information – touch (parietal lobe), vision (occipital lobe), hearing, and taste (insula and parietal lobe). Extensive regions of the cortex are classified as association areas, such as the prefrontal cortex and posterior parietal cortex, that are not solely sensory or motor in function.

The cerebellum, located in the hindbrain, is organized similarly to the cerebrum, consisting of a cortex and deep grey matter connected by its own specialized white matter tracts (42). Functionally, the cerebellum is recognized as an important motor control structure involved in balance and coordination and serving as an information relay center to the motor areas of the cortex. The cerebellum has also shown to be implicated in a number of non-motor cognitive functions such as attention and executive functions, that are mediated through specific cerebral-cerebellar loops (43).

The brain stem forms the stalk of the cerebrum and consists of the diencephalon (containing the thalamus and hypothalamus), the midbrain, the pons, and the medulla (**Figure 3B**). The functions housed deep in the brain within the brain stem are critical for survival and brain connectivity as this region serves as the primary conduit for white matter tracts (such as the corticospinal tract) connecting the cerebral cortex to the spinal cord and subcortical structures (40). The thalamus is a centralized relay center, providing most sensory and motor input to the cerebral cortex via specific nuclei. The hypothalamus is a vital control center for many basic bodily functions: regulating homeostasis, energy intake (governing the brain-body nutrient axis and food intake), sleep and circadian rhythms, temperature regulation, and fluid balance (44,45). It also contains memory circuitry that communicates directly with the hippocampus. The midbrain, pons, and medulla contain essential grey matter nuclei for motor control (both voluntary and involuntary), specialized sensory systems, and internal regulatory systems, including pain modulation (42).

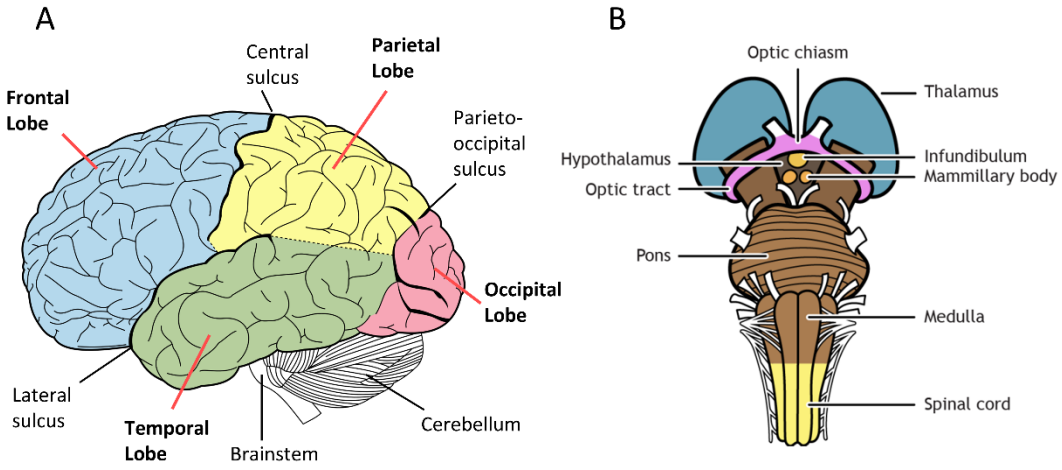


Figure 3. (A) Shows the cerebrum divided into four lobes separated by distinct fissures (or sulci). **(B)** Representation of the brain stem and its main features, including components of the midbrain, the pons, and medulla. Brainstem: by Casey Henley, licensed under a Creative Commons Attribution Non-Commercial Share-Alike (CC BY-NC-SA) 4.0 International License.

1.2.2 Blood supply of the brain

Due to its extensive connectivity the brain is highly energy-demanding, consuming approximately 20% of the body's total glucose and oxygen despite making up only 2% of body weight (46). The largest portion of this energy is spent on neuronal computation and information processing, with significantly more energy utilized in the grey matter than in the white matter (46). Specifically, processes related to synaptic transmission are estimated to be the main energy consumers, accounting for approximately 43% of the grey matter's energy budget (47). The brain's extensive vasculature assures there is constant brain perfusion to meet this extremely high energy demand.

The brain's blood supply is derived from four main arteries: the paired internal carotid arteries, which form the anterior circulation, and the paired vertebral arteries, which create the posterior circulation (48) (**Figure 4**). The vertebral arteries supply structures such as the cerebellum and brainstem before merging to form the basilar artery along the ventral surface of the brainstem. The basilar artery then bifurcates into the posterior cerebral arteries. At the base of the brain, the internal carotid arteries and their branches, the anterior cerebral arteries, connect with the posterior cerebral arteries through the anterior and posterior communicating arteries, forming the Circle of Willis (49). Traditionally, the Circle of Willis is thought to provide collateral (cross-hemisphere) circulation, serving as a compensatory mechanism that allows blood to bypass blockages in major vessels on one side of the brain via the communicating arteries (50). However, recent evidence indicates that these communicating arteries are often underdeveloped or absent in many individuals, which can hinder effective blood transfer and reduce this redundancy. An alternative hypothesis posits that the primary role of the Circle of Willis is to function as a passive pressure-dissipating system (50). This system helps protect the cerebral microvasculature and the blood-brain-barrier (BBB) from hemodynamic stress induced by asynchronous fluctuations in blood flow. Ischemic strokes often occur due to blockages in major input arteries, like the internal carotid artery, or in terminal branches, particularly within the territory of the middle cerebral artery, the most common site for strokes (51).

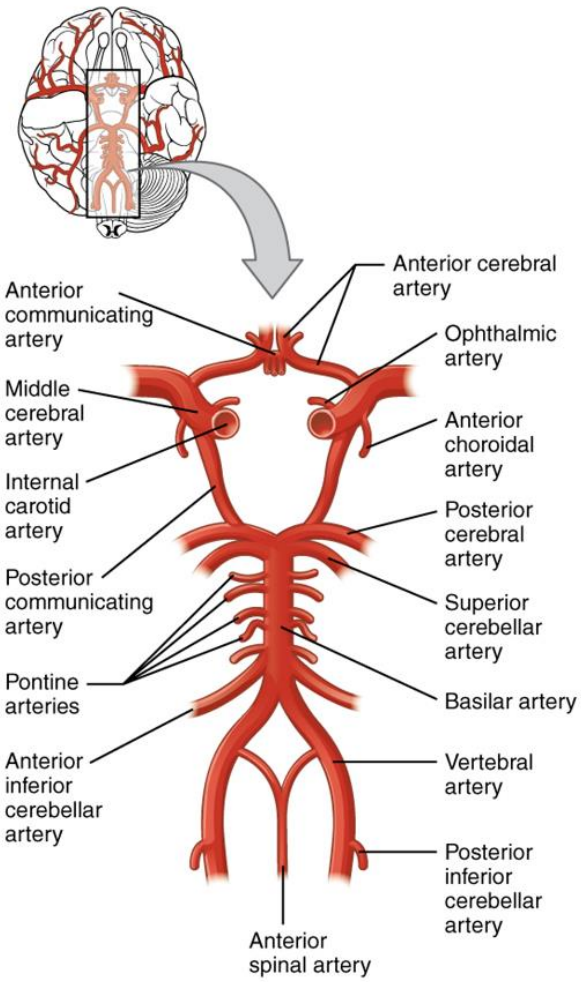


Figure 4. Schematic view of the main arteries supplying the brain giving rise to the circle of Willis. *Attribution: Chapter 11. Authored by: OpenStax College. Provided by: Rice University. Project: Anatomy & Physiology. License: CC BY: Attribution.*

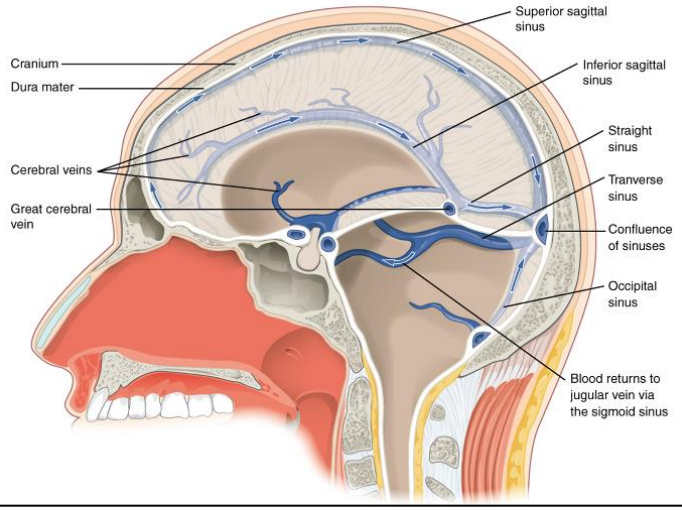


Figure 5. The dural and sinuses and veins showing that blood from the brain drains through a series of sinuses that connect to the jugular veins. *Attribution: Chapter 11. Authored by: OpenStax College. Provided by: Rice University. Project: Anatomy & Physiology. License: CC BY: Attribution.*

The brain's venous system is composed of a network of venous sinuses that collect deoxygenated blood and metabolic waste from the brain parenchyma, meninges, eyes, and cerebrospinal fluid (CSF) that ultimately converge toward the internal jugular veins (**Figure 5**) (52). Although venous pathology is relatively uncommon compared to arterial disease in the brain, the absence of valves allows for free communication with veins outside the brain. This can potentially lead to the spread of infectious pathogens into the brain. Additionally, dysfunction within this system can result in serious conditions such as cerebral venous thrombosis or subdural hematoma (52).

1.2.3 The Neurovascular Unit

Nutrients arriving to the brain through arteries and are exchanged with neurons and glial cells via the neurovascular unit (NVU). The NVU defines the essential, intricate relationship between brain cells and the cerebral vasculature. It is composed of vascular cells including endothelial cells and mural cells (such as smooth muscle cells and pericytes) found along the blood vessel walls, in addition to neurons, interneurons, and glial cells like astrocytes, microglia, and oligodendrocytes (**Figure 6**) (53). Additionally, a basement membrane formed by brain endothelial cells and the extracellular matrix supports these components. The NVU components together coordinate signals across the vascular network to respond to the brain's energy demands by regulating cerebral blood flow in a process called "neurovascular coupling" (53). It is hypothesized that neurovascular coupling is initiated through signaling at the microvascular level in capillaries and then propagated upstream. Specifically, this regulation is initiated by neurons and transmitted through astrocytes end-feet to mural cells which constrict or dilate vessels (53). The NVU is also responsible for maintaining the integrity of the BBB, a core component of the NVU, formed by brain endothelial cells interconnected by impermeable tight-junctions. The BBB regulates the passage of molecules between circulation and the brain parenchyma which is essential for maintaining homeostasis of the internal environment in the brain (54). Its integrity and transport functions are supported by pericytes, astrocytes, and the basement membrane (54). A compromised NVU results in reduced

cerebral blood flow and BBB dysfunction which are hallmarks of cSVD and has also been extensively linked to neurodegeneration, in particular through hampered clearance of amyloid-β (Aβ) (53,55).

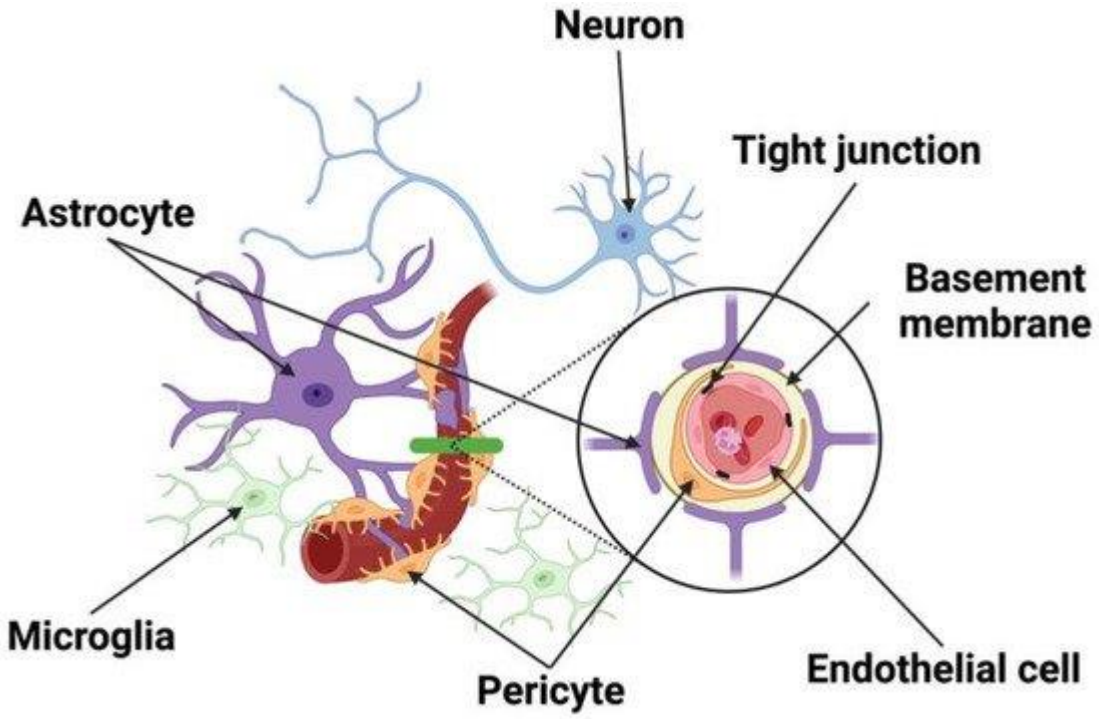


Figure 6. The Neurovascular Unit. Endothelial cells forming the blood vessels are encased by a basal lamina/basement membrane and are bound by tight junction proteins. Located in the brain parenchyma, astrocytes make contact with both pericytes and endothelial cells at the capillary wall, while pericytes are situated between the end feet of astrocytes and endothelial cells. *Adapted from Sato et al. (2022)* (56)

1.3 The spectrum of cerebrovascular disease

The spectrum of vascular brain health and cerebrovascular pathologies can be effectively understood through the metaphor of an iceberg (**Figure 7**). Historically, clinical stroke was the first widely examined measure of vascular brain health, but it represents only the visible tip of the iceberg of vascular brain injury (57). In addition to clinical stroke, the term vascular cognitive impairment (VCI) was introduced to describe the entire range of vascular brain pathologies that impact cognitive function, from subjective cognitive decline and mild cognitive impairment (MCI) to severe dementia (58). In fact, it is increasingly understood that most dementia cases are caused by a mix of neurodegenerative and vascular pathology (4,5,58). In some cases, more subtle and gradual symptoms of VCI can be detected early on, such as mental slowness, problems with executive function, memory issues, gait disturbances, apathy, anxiety, and depression (58). Yet, most vascular brain injury lies below the surface in the form of covert cSVD, characterized by pathological changes in the structure and function of small brain vessels and capillaries without clinical symptoms (59). In this section, we start with the foundation of cerebrovascular disease,

cSVD, and then discuss the clinical presentation, epidemiology, risk factors, and pathophysiology of stroke and dementia.

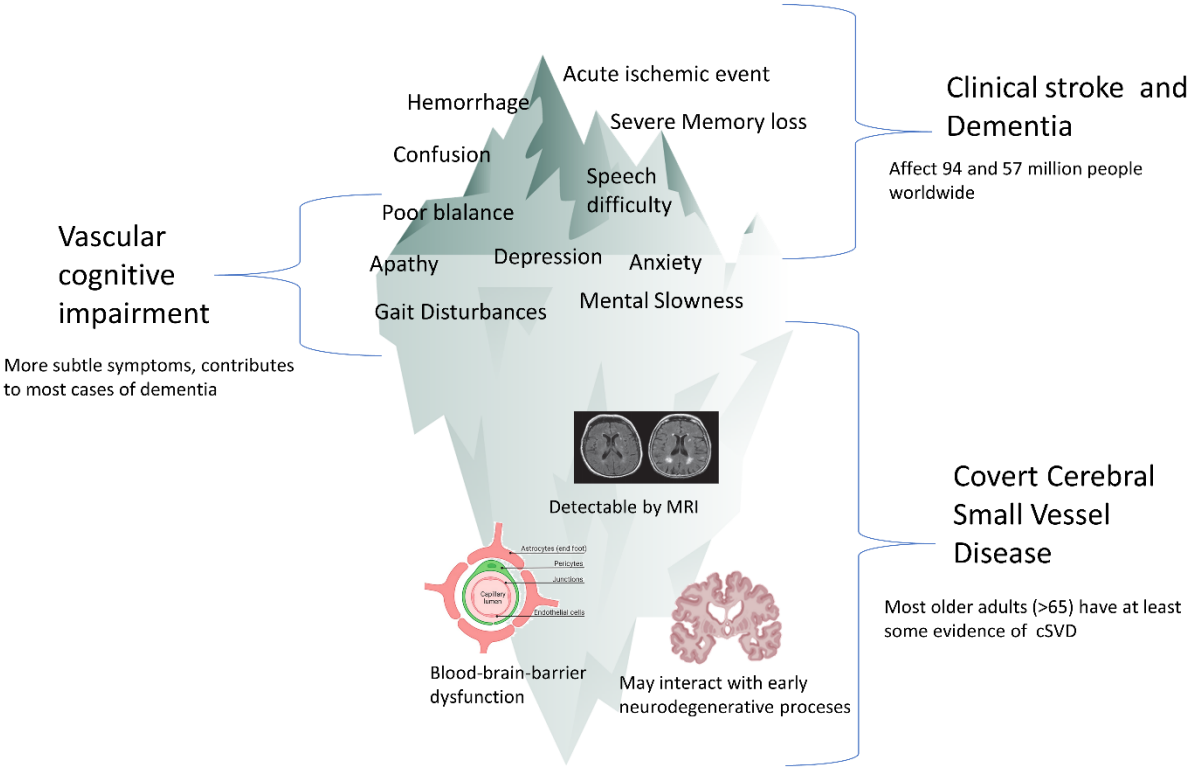


Figure 7. Iceberg model to describe the spectrum of cerebrovascular disease. At the tip of the iceberg is clinical stroke and severe vascular dementia. Just below the surface we may find more subtle symptomatology in the form of mild cognitive deficits, gait disturbances, and depression. The majority of cerebrovascular disease manifests sub-clinically, well below the surface, in the form of cSVD, detectable through brain imaging, that is additive or may interact with other neurodegenerative processes.

1.3.1 Cerebral Small Vessel Disease

Clinical cSVD: cSVD refers generally to pathological processes that affect the small vessels of the brain with focus on small arteries, arterioles and capillaries (60). Clinically, cSVD is thought to be responsible for approximately 25% of ischemic strokes in the form of lacunar stroke resulting from small, deep infarcts (60). cSVD also causes the majority of intracerebral hemorrhages, and is the most common pathology underlying VCI (7). Moreover, cSVD lesions commonly coexist with AD neurodegenerative pathologies providing what is likely an independent, additive contribution to dementia that lowers the threshold for clinical AD diagnosis (58). Overall, it is estimated that cSVD contributes to about 50% of dementias worldwide (61). Other clinical features are also recognized such as motor impairment, impaired balance, falls, and behavioural symptoms like depression, apathy and personality changes (62). However, most vascular brain injury due to cSVD that is 'covert' or 'silent', in that many affected people do not exhibit symptoms. Therefore, 'covert' cSVD is often studied as an endophenotype and major risk factor predisposing to stroke and dementia.

Imaging markers of cSVD: Most often, 'covert' cSVD manifests as vascular lesions traditionally detectable on brain MRI (57,59). MRI-markers of cSVD have shown to be associated with cognitive impairment, dementia, depression, mobility issues, and increased risk of stroke and worse outcomes post-stroke (63,64). The most routinely recognized MRI-markers, which are standardized according to the 2nd interation of the standards for reporting vascular changes on neuroimaging (STRIVE-2) criteria, are white matter hyperintensities (WMH), perivascular spaces (PVS), lacunes of vascular origin, small subcortical infarcts, cerebral microbleeds (CMB), and brain atrophy (65). Some of these markers are shown in **Figure 8** and summarized in **Table 1**.

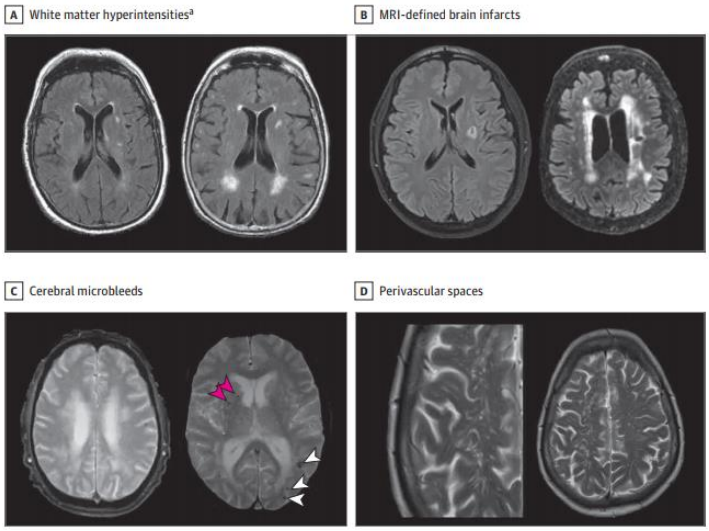


Figure 8. MRI markers of covert cSVD from *Debette and Markus (2019)*. (A) White matter hyperintensities (B) Silent brain infarcts both with (right) and without (left) WMH. (C) Single cerebral microbleed (left) compared to multiple CMB (right). (D) Perivascular spaces following the shape of deep penetrating arteries.

MRI Marker	Description (Imaging/Pathology)
White Matter Hyperintensities (WMH) (of presumed vascular origin)	Appear as diffuse areas of high signal intensity (hyperintense) on T2-weighted or FLAIR sequences (7). Periventricular WMHs are often symmetrical around ventricles, while deep WMHs are smaller and sometimes asymmetrical (59). They represent areas of tissue damage (gliosis, axonal loss, and demyelination) (63).
Silent Brain Infarcts	These are areas of infarction, typically small and subcortical, which are asymptomatic ("covert"). Defined by neuroimaging evidence of acute ischemia in the territory of a single perforating artery, typically defined as > 3mm and ≤20 mm in axial section, though some studies consider a 15mm threshold (66).
Lacunes (of presumed vascular origin)	Round or ovoid, subcortical, **fluid-filled cavities** 3–15 mm in diameter. They are the resulting cavities after the healing of lacunar infarcts or small hemorrhages (59).
Cerebral Microbleeds (CMB)	Small (2–10 mm) areas of low signal ("signal void" or "black spots") seen prominently on T2-weighted or susceptibility weighted sequences (59,63). They indicate perivascular deposits of hemosiderin from prior subtle hemorrhages (67).
Enlarged Perivascular Spaces (PVS)	Appear as small, fluid-filled structures, that run along the outside of penetrating blood vessels. They are typically visible with heavily T2-weighted MRI sequences, and exhibit a signal intensity similar to CSF on all sequences. They are predominantly found in the basal ganglia, in white matter (e.g, centrum semiovale), and also in the hippocampus (68).

Table 1. Description of MRI markers of cSVD.

These cSVD-MRI markers frequently co-exist, are associated with each other, and related to similar clinical outcomes (65,68). However they may still represent distinct disease processes reflected in cSVD heterogeneity and be more indicative of ischemic (e.g., brain infarcts and lacunes) versus hemorrhagic risk (e.g., CMBs) (69). Efforts have been made to develop a comprehensive phenotype to study cSVD. STRIVE-2 proposes a 'summary SVD score' that includes WMH, lacunes, PVS, and CMB, but these scales are still under development and their most useful applications remain unclear (65). Additionally, an extreme cSVD phenotype has been used for research purposes, constructed based on extremes of WMH distribution and presence or absence of lacunes to define two risk groups representing extensive or minimal cSVD burden (70).

Beyond structural MRI, diffusion-MRI (dMRI) metrics may better capture cSVD burden and offer insights into white matter microstructure (71). Diffusion tensor imaging (DTI) markers such as fractional anisotropy, mean diffusivity, and a composite of the two, peak skeletonized mean diffusivity (PSMD), have been proposed as emerging markers (72,73). Recently, Neurite Orientation Dispersion and Density Imaging (NODDI) markers derived from multi-shell diffusion-MRI techniques, show promise beyond DTI to describe underlying tissue properties, allowing for even more refined characterization of the white matter microstructure (71,74). Further, NODDI metrics may capture cSVD-related microstructural changes early in life, and allow for further regional and biological specificity (71).

Pathology: cSVD is understood as a heterogenous disorder with multiple causes. First, there are monogenic or familial causes of cSVD, including cerebral arteriopathy with subcortical infarcts and leukoencephalopathy that can be either autosomal dominant (CADASIL) or recessive (CARASIL). These monogenic forms of cSVD are described more in the next chapter. More common is the sporadic form of cSVD which can be broadly classified into two major categories: arteriolosclerosis and cerebral amyloid angiopathy (CAA). Arteriosclerotic cSVD is characterized by the thickening of vessel walls due to accumulation of fibro-hyaline material and collagens in the vessel walls, and the loss of smooth muscle cells (59). CAA, alternatively, is a distinct pathology involving the accumulation of Aβ protein in the walls of small vessels, particularly leptomeningeal and cortical vessels, and is specifically associated to hemorrhagic, such as CMBs and more clinically, lobar ICH (59,61,75). It is also particularly prevalent in the elderly (75).

Regardless of the specific cause, damage to the microvasculature compromises the structural integrity and function of these vessels leading to the lesions we see on MRI. The convergent pathogenic hallmarks of cSVD include reduced cerebral blood flow, vascular endothelial dysfunction, BBB dysregulation, vascular inflammation, and impaired brain waste drainage

(59,75). At the core of cSVD is dysregulation of the NVU which precipitates BBB leakage, allowing increased fluids and proteins to leak into the brain's extracellular space. This causes further glial cell dysfunction resulting in impaired myelination and repair, microglial-driven inflammation, and damage to astrocyte end-feet (61). However, the precise underlying pathology and mechanisms of cSVD remains largely unclear due to its clinical and biological heterogeneity and complexity.

Risk Factors: Age is the strongest risk factor for cSVD, with prevalence of WMH affecting up to 90% of individuals by age 65 (75).The most predominant modifiable risk factor for cSVD is hypertension, especially for the atherosclerotic-type (75). Other well-known metabolic and lifestyle factors, including, type-2 diabetes, smoking, and dyslipidemia have also been shown to be associated with cSVD MRI-markers, though some of these associations seem to be study and MRI-marker dependent (59). However, most of the variance of cSVD burden remains unaccounted for. In one study, vascular risk factors taken together only explained about 2% of the variance in WMH (76). Increasing evidence also points to a life-course model where factors present early in life influence later-life cSVD risk (77). These include midlife vascular risk factors in addition to factors rooted in prenatal development and childhood such as prenatal nutrition, early life socioeconomic status, education, and cognitive ability (77).

Prevention and treatment: The primary approach to prevent covert cSVD, according to the *European Stroke Organization (ESO)* guidelines, is to control hypertension (9). Additionally, lifestyle modifications focusing on vascular health, such as regular physical activity, smoking cessation, a healthy diet, and good sleep habits, are also strongly recommended (9).

There is mixed evidence for pharmaceutical interventions which typically aim to control for vascular risk factors (78). Antihypertensive treatment, on one hand, is strongly recommended to prevent progression of cSVD lesions and reduce subsequent stroke and dementia risk (9). The SPRINT trial showed that intensive blood pressure lowering (targeting systolic blood pressure (SBP) <120 mmHg) significantly reduced the progression of WMH and lowered the combined rate of MCI or probable dementia (79–81). Lipid-lowering agents such as statins may also be beneficial for treating cSVD due to their effectiveness to manage atherosclerosis and reduce stroke risk overall, though high quality evidence to prevent clinical outcomes in cSVD is lacking (9). Antiplatelet and anticoagulation therapy is cautioned in cSVD patients due to the increased risk of hemorrhage (especially due to CAA). The *ESO* guidelines recommend against the use of single or dual therapy with aspirin and/or clopidogrel unless there is sufficient indication otherwise (9). Overall, beyond typical lifestyle modifications and pharmaceutical interventions to control for

known vascular risk factors, there are no disease modifying therapies or mechanism-based treatments readily available for cSVD. This highlights the need a greater understanding of cSVD aetiology, and strategies to identify individuals at increased cSVD risk that emphasize a life-course perspective.

1.3.2 Stroke

Clinical definition: Stroke is defined by the World Health Organization (WHO) as rapidly developing clinical signs of focal disturbance of cerebral function lasting more than 24 hours or leading to death with no other explanation other than a cerebrovascular cause (82,83).

Sub-types: Stroke is primarily categorized as ischemic or hemorrhagic. Ischemic strokes result from an acute arterial occlusion, or infarct, blocking the blood supply to part of the brain. Ischemic strokes can be further classified into specific sub-types based on primarily their underlying cause or presumed etiology, and secondarily by the location of the infarct or clinical presentation (83). The most common, and long-standing, classification scheme is called TOAST (84) which includes four major categories:

- **Large artery atherosclerosis (LAS)** resulting from artery-to-artery thromboembolism originating from atherosclerotic lesions, or plaques, of the internal carotid artery or major intracranial arteries.
- **Cardioembolism (CES)** due to an embolism, or clot, originating from the heart. A key diagnostic indicator of this type is atrial fibrillation (AF), a heart rhythm disorder where the atria of the heart beat in a rapid, irregular, and uncoordinated way.
- **Small vessel disease or lacunar infarction (SVS)** where there is occlusion of the lenticulostriate arteries, deep, small blood vessels branching from the middle cerebral artery, or vertebrobasilar perforators, tiny arteries that arise from the vertebral or basilar artery.
- **Other causes** such as cervical artery dissection, vasculitis, or hereditary conditions.
- **Underdetermined causes** which are also called cryptogenic stroke.

The INTERSTROKE study, based on the TOAST criteria, found that among 2,337 ischemic stroke cases with full diagnostic work-up, that SVS accounted for 44% (n=1039), LAS for 19% (n=437), CES for 9% (n=220), other causes 5% (n=124), and 22% were of undetermined cause (n=517) (85). These numbers vary depending on the study.

Hemorrhagic strokes occur when there is a bleeding in or around the brain that is not due to trauma, and are typically more catastrophic than ischemic strokes. Bleeding from a blood vessel within the

brain is termed an intracerebral hemorrhage (ICH), while bleeding into the subarachnoid space surrounding the brain is called a sub-arachnoid hemorrhage (SAH). Roughly 80% of ICH are caused by sporadic cSVD either in the form of deep perforating artery vasculopathy or CAA (86). Other causes include malformations in the brain macro-vasculature, and clotting in cerebral veins leading to hemorrhagic transformation (83).

Global burden: Between 1990 and 2019, the global incidence of stroke increased by 70% and the prevalence by 85%. Ischemic stroke account for an estimated 62.4% of all incident strokes, ICH constituted 27.9%, and SAH represented 9.7% (1). In high-income countries, the yearly incidence of ischemic stroke is about 40 per 100,000 people and 15 per 100,000 people for ICH, while in LMICs, these numbers increase to 150 and 97 per 100,000 people respectively (1). Indeed, ICH, in particular, accounts for only 15% of all strokes in high-income countries, but 30% of strokes in LMICs (83). As highlighted in the first section of this chapter, this increase is partly driven by population growth and ageing (1). However, we have also seen a substantial 50% increase in the incidence of ischemic stroke in young- to middle-aged adults (18-50 years old) over the last decade (87). In terms of sex differences, the absolute global burden was greater in males, but incidence and prevalence of stroke was greater in females (1).

Epidemiology and risk factors: The Global Burden of Disease study estimated that almost 90% of global burden of stroke was attributable to 19 potentially modifiable risk factors (1). This estimation was similar across both ischemic and hemorrhagic sub-types. The risk factors for stroke can clustered into metabolic factors, behavioural/lifestyle factors, and environmental factors, attributable to an estimated 71%, 47%, and 38% of global stroke burden respectively (1). Metabolic factors, in order of contribution to stroke burden, include high blood pressure, body mass index (BMI), fasting blood glucose (i.e., diabetes), LDL cholesterol, and kidney dysfunction. Behavioural and lifestyle risk factors include smoking, unhealthy diets, (including high alcohol consumption) and low physical activity. Environmental factors include ambient particulate matter ($PM_{2.5}$), household air population from solid fuels, extreme ambient temperatures, and lead exposure. Importantly, many of these factors are interrelated. Many of the major risk factors have much higher prevalence in LMICs (84). Notably, uncontrolled hypertension, diabetes, obesity, smoking, poor diet and lack of exercise are exacerbated by reduced access to primary care (83).

Stroke prevention: Primary prevention of stroke mainly involves lifestyle modification and aggressive management of vascular risk factors, especially at mid-life (88,89). Lifestyle modifications are foundational for primary prevention, including a healthy diet (e.g., the Mediterranean diet), weight loss, cessation of smoking, and regular physical activity (88).

Additionally, reduction in salt intake has been shown to reduce cardiovascular mortality in hypertensive patients, and specific nutritional supplementation, like folic acid or B-vitamin complexes, may also be helpful (88). Control of hypertension is largely considered the most effective preventive strategy for stroke, targeting <140/90 mm Hg. Moreover, every 10 mmHg reduction in systolic blood pressure and 5 mmHg reduction in diastolic blood pressure has shown to reduce the risk of stroke by an estimated 41% (88,90). Lipid management with statins and/or PCSK9 inhibitors has also shown to be effective for primary stroke prevention (88). To prevent CES in particular, management of AF is critical. This primarily includes treatment with anticoagulants which can reduce risk of stroke by 60-80%, though anticoagulation increases the risk of bleeding events (88). Direct oral anticoagulants (DOAC) are now considered frontline therapy over Vitamin-K antagonists (e.g., warfarin) as there is a substantial reduction in comparative risk of hemorrhagic stroke (91). AF patients may also be treated with antiplatelet therapies (e.g., aspirin or clopidogrel), although these have shown to be less effective than DOACs while having similar bleeding risks (88).

Secondary prevention of stroke is also of utmost importance as about 20-30% of strokes are recurrent (92). The risk of recurrent ischemic stroke is about 15% at 1-month, and in the long-term, reaches 25% at 5-years and 40% at 10-years (88,92). However, potentially up to 80% of recurrent strokes are preventable with proper care management (88). Treatment course depends largely on the underlying stroke pathology, making correct classification of stroke a vital component of secondary prevention (93). Most secondary stroke prevention strategies are similar to primary stroke prevention. For ischemic stroke, this includes aggressive management of hypertension, use of DOACs in patients with AF, and long-term antiplatelet therapy (88,93,94). Some more invasive procedures may be relevant for some patients, including for example, carotid revascularization for patients with stenosis, or patent foramen ovale closure in younger patients with cryptogenic (undetermined) stroke (88,93).

Though less prevalent, ICH is a more devastating stroke than ischemic stroke with an estimated 35% fatality rate within 1-month compared to about 15% for ischemic stroke (95,96). Data from the Dijon stroke registry indicated that the 1-year relative survival for ischemic stroke was 82% compared to 56% for ICH (97). Recurrence rates of ICH are estimated to be between 1.2-3% per year, with significantly higher recurrence for those with CAA (7.1%) compared to non-CAA-related ICH (1.1%) (98). ICH survivors are also at high risk of subsequent ischemic events, with ICH location and comorbid AF being key factors in determining a patient's risk profile (99). For ICH secondary prevention, again, similar recommendations are made for primary stroke

prevention (88). In ICH survivors the main clinical dilemma is to balance competing risks of thrombosis and recurrent bleeding, particularly with regards to resumption or initiation of anticoagulation therapy (100). In patients with higher risk of ischemic stroke, antiplatelet therapy can be resumed without increasing the risk of ICH according to the results of the RESTART trial (101). However, there remains a lack of treatment guidelines in particular for individuals who have survived an ICH who also have AF. Around 25% of ICH survivors also have AF, and randomized trials investigating resumption of anticoagulation remain inconclusive (100–103). There is also evidence lacking with respect to lipid-lowering therapies, such as statins (100). Alternatively, left atrial appendage closure may be effective to reduce thrombosis risk among those with contraindications for anticoagulation (88,104). Overall, other than management of hypertension and lifestyle modifications, current evidence is lacking for identifying the proper course of treatment for secondary prevention in ICH survivors and dedicated prospective studies and randomized trials are needed.

1.3.3 Dementia

Clinical Presentation: Dementia is a clinical syndrome that is marked by cognitive impairment severe enough to impede independence and activities of daily living. Onset of dementia is gradual with a long prodromal, or pre-clinical stage, that begins symptomatically with MCI in a single or multiple cognitive domains, but where functional capacities remain largely intact (105). Dementia is often diagnosed in the clinic using the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-V), where it is also referred to major neurocognitive disorder (106,107). The core elements of the clinical definition include:

- **Cognitive decline and impairment:** There is evidence of substantial impairment in at least one or (usually) multiple cognitive domains, such as, attention, executive function, language, learning, memory, perceptual-motor, or social cognition (106). These cognitive domains are preferably documented by standardized neuropsychological tests. The impairment also must importantly represent a decline from a previously higher level. Cognitive impairment may additionally include changes to mood and behaviour.
- **Interference with daily function:** Cognition is sufficiently impaired where it impacts independence in everyday functioning, such that it significantly interferes with work or social activities or relationships with others (106). This criteria distinguishes dementia from MCI.

It also must not be explained by cognitive deficits that occur during episodes of delirium or by other mental disorders. Dementia assessment is typically performed when the patients themselves and/or their family members express concern about changes in cognitive function (108).

Dementia sub-types: The definite classification of dementia is based on the underlying neuropathology determined by autopsy, but can also be partly distinguished based on clinical features and symptomatic progression (109). The main aetiologies of dementia are Alzheimer's Disease (AD), vascular dementia (VaD), frontotemporal dementia (FTD), and Lewy body dementia (LBD) including Parkinson's disease dementia (PDD) (**Table 2**

Dementia Sub-type (% of dementia cases)	*Main neuropathology*	*Clinical symptoms*
Alzheimer's Disease (~60%)	A dual proteinopathy defined by the coexistence of extracellular aggregates of Aβ-42 fibrils (amyloid plaques) and intracellular aggregates of hyperphosphorylated tau (P-tau), termed neurofibrillary tangles (NFTs).	AD often involves prominent early episodic memory deficits and amnesia reflecting neurodegeneration in the temporal and parietal lobes (109). Individuals may also present decreased semantic fluency, difficulty with mathematical calculations (acalculia), and/or visuospatial deficits.
Vascular dementia (~15%)	Dementia symptoms may develop post-stroke or be due to progressive cSVD. In particular, there is a clear temporal relationship with a cerebrovascular event or presence of diffuse cerebrovascular disease on neuroimaging or neuropathogy (110). It frequently coexists with AD and LBD pathology (mixed pathology) (109).	Most heterogenous presentation. Early changes often include reduced cognitive flexibility and verbal memory retrieval. Frontal lobe deficits are common including executive dysfunction, poor attention and concentration, and increased irritability (109).
Frontotemporal dementia (~2%)	Arises from frontotemporal lobar degeneration, denoting the degeneration of cortical and subcortical structures in the frontal and temporal regions of the brain. Frontotemporal lobar degeneration is caused predominantly by intracellular aggregates of tau, TAR DNA-binding protein 43 (TDP-43) or fused in sarcoma (FUS) (109).	Typically, an early-onset dementia (<65 years old) (111). Most often characterized by changes in behaviour and reduced language and social skills with relative sparing of memory and visuospatial functions early in the disease course (109).
Lewy body dementias (~10%)	Defined by characteristic intracellular α-synuclein aggregates (Lewy bodies). LBD encompasses Dementia with Lewy Bodies and Parkinson Disease Dementia. Frequently, patients with LBD show mixed pathology, often involving vascular disease and/or AD pathology (112).	Manifests symptomatically between the ages of 60 and 90 years**.** LBD core features include fluctuating cognition, recurrent visual hallucinations, sleep disturbances and parkinsonism. PDD is characterized by cognitive decline after onset of parkinsonism with executive dysfunction (113).

Table 2. Overview of dementia aetiologies. Most of the pathological and clinical symptoms are described in Elahi and Miller (2017) (109). Sub-type classification %'s are taken from the UK Dementia Statistics Hub operated by Alzheimer's Research UK based on their national health data registry (114).

AD is characterized by extracellular accumulation of extracellular plaques containing Aβ and tau-containing intracellular neurofibrillary tangles (NFTs) (115). Age is the strongest risk factor for AD, and most cases are sporadic, 'late-onset', after 65 years old, while early onset AD are typically

caused by rare genetic mutations (115). Originally, individuals were diagnosed with AD if they had amnestic dementia and other causes could be ruled out, however, integration of clinical symptoms with objective biomarker evidence defines AD as a neurobiological condition (116). The diagnostic process begins with clinical assessment for the presence and severity of cognitive impairment, and consideration of other causes. AD-specific biomarkers have helped to establish levels of certainty for AD pathology that were previously impossible apart from autopsy. To this end, amyloid and tau biomarkers along with non-specific biomarkers of neurodegeneration have been combined in the widely used "A-T-N" framework intended to classify individuals on the AD spectrum (116).

- **A (Amyloid)** status is determined by Aβ-positon emission topography (PET) imaging showing Aβ plaque accumulation or measuring decreased Aβ-42 (or the Aβ42/Aβ40 ratio) in CSF.
- **T (Tau)** status is determined by Tau-PET imaging showing tauopathy distribution or measuring increased p-tau181 in CSF.
- **N (Neurodegeneration)** markers include FDG-PET to determine patterns of temporal-parietal hypometabolism and structural MRI to assesses regional atrophy.

Together, they allow for more precise assessment of the presence and severity of AD pathology. Notably, the amyloid cascade hypothesis, positing that the accumulation of Aβ was the main cause of AD, has dominated AD research (117). However, many elderly individuals have elevated brain Aβ levels but are cognitively normal (118). After many clinical trial failures in the past 25 years new avenues are being explored to treat and prevent AD (117). Alternative AD models put *APOE*, the strongest genetic risk factor, at the forefront suggesting AD pathology may be stratified based on genetic background (119). These models also put more weight on the contribution of neuroinflammatory, environmental, metabolic and vascular factors that are independent of, or potentially interact with, Aβ (119,120).

Pure VaD alone probably accounts for only ≤10% of dementia cases (5), yet, vascular pathologies, contributing to VCI, are found in up to 75% of patients with dementia (**Figure 9**) (58,121). Stroke is the most apparent driver of VCI, and post-stroke MCI and dementia is estimated to occur in 38% and 18% of individuals respectively (122–124). As mentioned, cSVD is the most common substrate for VCI, and can cause cognitive decline in the absence of clinical stroke. Indeed, mixed dementia pathology, most frequently a mix of vascular and AD-related processes, is very common, especially in the elderly (5,125). A recent study pooling data from six community-based autopsy studies in people over 80 years old who died with dementia showed that 91% of participants had evidence for more than one neuropathology (126). One study found that AD and vascular pathologies present in autopsies were largely uncorrelated, suggesting that while they might have independent causes, they could interact to accelerate and promote cognitive decline and dementia (126).

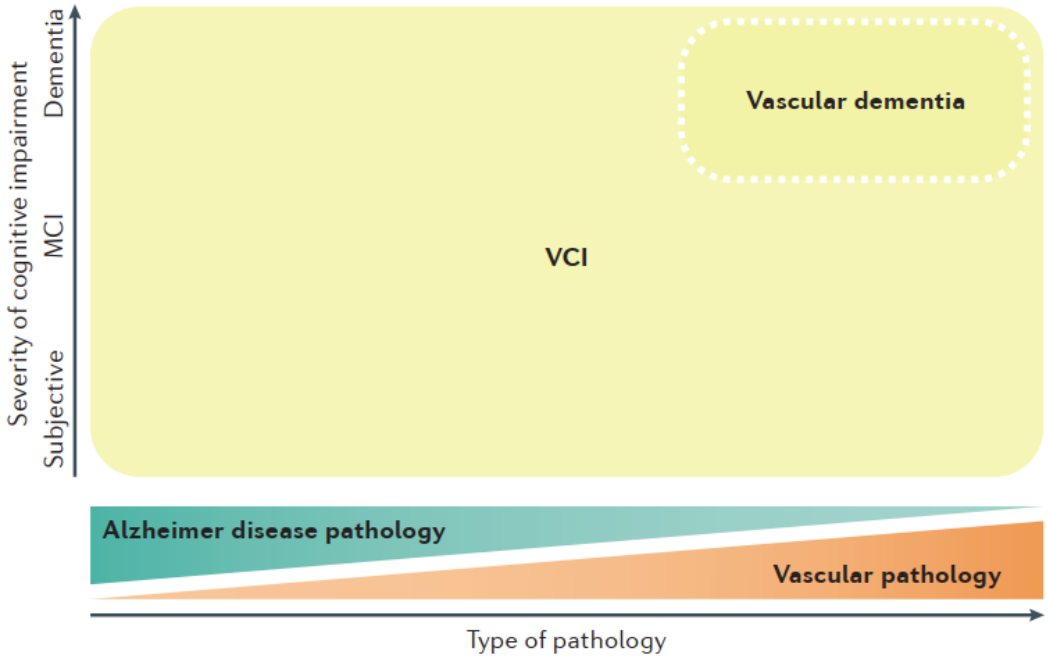


Figure 9. Conceptual relationship between AD, VCI and VaD. Dementia pathology is increasingly understood to often arise from a mix of vascular and AD pathology. VCI is an umbrella term associated with any severity of cognitive impairment arising from vascular pathology. Individuals with AD likely also have some co-occuring vascular contributions to cognitive impairment as well. VaD is a subgroup of patients who have clinical dementia attributable largely to vascular cases. Adapted from van der Flier et al. (2018) (58)

Global burden: The absolute number of people affected by dementia globally increased by approximately 117% between 1990 and 2016 (127). In 2019, approximately 57.4 million people worldwide were living with dementia (95% CI 50.4–65.1 million), and this number is projected to increase to 152.8 million (130.8–175.9 million) by 2050 (23). Additionally, women consistently showed higher prevalence rates than men (female-to-male ratio = 1.70), a trend expected to continue to 2050 as well (23). Part of the sex differences could be explained by higher life expectancy among women, but additional inequities could also be driving finding (23). In both

sexes, dementia prevalence roughly doubles every five years up to age 85 (23). Despite the expected increase in overall cases, the age and sex standardized global prevalence is expected to remain relatively stable, with only a slight 0.1% change from 2019 to 2050 (23). Data from high-income countries shows that in fact age-specific dementia rates have decreased over the past two decades (128). However, increasing prevalence in preventable vascular risk factors (e.g., obesity and diabetes) may cause age-specific rates to increase in the future (129).

Regionally, population growth is the primary driver in sub-Saharan Africa, while aging predominates in East Asia. The smallest increases are anticipated in high-income regions, such as high-income areas of the Asia Pacific (53%) and Western Europe (74%), while the most significant growth is projected in North Africa and the Middle East (367%) and eastern sub-Saharan Africa (357%). Overall, regions with low Socio-Demographic Index (SDI) quintiles are expected to see the highest percentage increases, with the lowest SDI group anticipating a 330% rise in total cases by 2050 compared to 113% for those in the high-middle SDI quintile.

In terms of MCI, a meta-analysis of 66 studies estimated the overall prevalence at 15.56% worldwide in community-dwelling adults aged 50 years and older (130), and about 30%, nearly double, in patients with hypertension (131).

Risk factors: The 2024 *Lancet Commission* on dementia estimated that, from a life-course perspective, about 45% of dementias could be theoretically prevented by eliminating 14 modifiable risk factors (**Figure 10**) (129). Many mid-life risk factors are congruent with risk factors for stroke, including hypertension, obesity, smoking, diabetes, physical inactivity, excessive alcohol consumption,

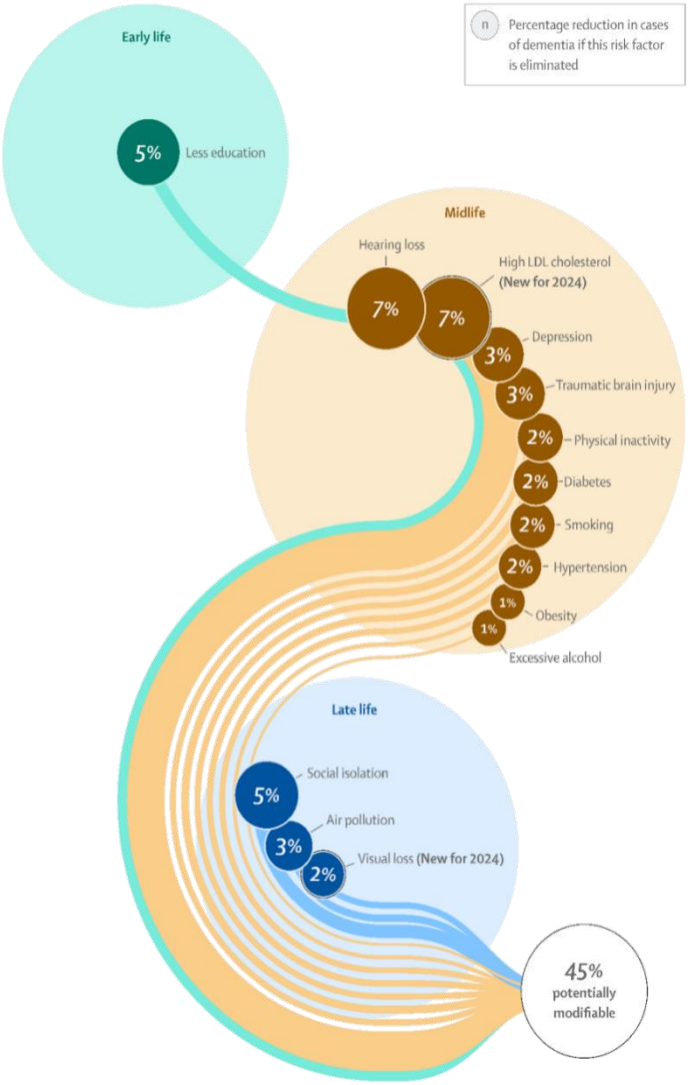


Figure 10. Lancet Commission figure for the 14 modifiable risk factors of dementia broken down in a life-course model. The % for each risk factor indicates the relative reduction in number of dementia cases if this risk factor were completely eliminated. *Adapted from Livingston et al. (2024)*

and high LDL cholesterol. Other risk factors, related to the disadvantageous social, environmental or economic circumstances, are in turn linked to other health factors that are associated with cognitive decline and dementia (132). Addressing these modifiable risk factors could potentially prevent or delay dementia onset by decreasing underlying vascular damage, reducing neuropathology, limiting stress and inflammation, and building cognitive reserve (i.e., the ability to withstand or resist neuropathologies before onset of symptoms) (129). Other potential risk factors for dementia include infectious diseases, with emerging evidence for increased dementia risk for those infected with diseases outside the CNS (e.g., herpes simplex) that likely act through systemic inflammation and long-term immune activation (133). Longitudinal evidence is also needed to determine the impact of COVID-19 on dementia risk.

The most significant non-modifiable risk factor for dementia is age. Genetics—the topic of the next chapter—also plays a pivotal role in dementia risk. In particular, the ε4 allele of apolipoprotein E (*APOE*) gene is the most well-established common genetic risk factor for AD. Biological sex is also an important modulator of the impact of risk factors (both modifiable and non-modifiable) on dementia risk (134), and there is mounting evidence for heterogeneity in dementia pathology between the sexes as well (134).

Primary and secondary prevention: The primary prevention of dementia, like stroke, also involves lifestyle modifications and management of vascular risk factors. Furthermore, greater cognitive and physical reserve across the life course and reductions in vascular disease have likely contributed to reduced rates of age-related dementia incidence in high-income countries (129). Additionally, modifiable risk factors can be categorized based on the life stage when intervention is most impactful with most of them being midlife (**Figure 10**) (131). Multidomain interventions, including a substantial non-pharmaceutical component, have shown consistent effectiveness for reducing dementia risk, and could be cost-effective in the long-term (135,136). For example, the FINGER study (137), which delivered lifestyle interventions and vascular risk monitoring to at-risk elderly people over two years demonstrated particularly encouraging results (138). Supporting the need for precision medicine approaches, FINGER and another multidomain trial, MAPT (141), had greater effectiveness in high-risk groups that were identified based on genetic, biological or clinical features. Such features included *APOE-ε4* carrier status (140), brain β-amyloidosis (141), and validated clinical risk scores (142,143). These prevention strategies included a physical fitness regime with a focus on aerobic exercise, cognitive training or stimulation, brain-health promoting diets (e.g., the Mediterranean diet) with supplementation of omega-3 polyunsaturated fatty acids

(i.e., fish oils), and vascular risk monitoring especially for management of blood pressure and diabetes (137,139).

Due to dementia's long prodromal period, secondary prevention focuses on interventions in individuals who show signs of MCI, have subjective cognitive complaints, or, more recently, have evidence of preclinical pathology determined by biomarkers or genetic risk (142). Secondary prevention occurs in the memory clinic, outpatient facilities in the context of neurology, geriatric, or psychiatric services, where patients undergo diagnostic workup, and are administered interventions that may include pharmaceutical drugs or selection into randomized trials for potential disease modifying therapies (142).

Currently pharmaceutical interventions, such as cholinesterase inhibitors and memantine for AD and LBD, show modest benefits in delaying symptom onset and progression (144,145). Other drugs target Aβ with monoclonal antibodies. While trials for Aβ-modifying therapies have been moderately effective for improving cognitive outcomes, they come with risk of amyloid-related imaging abnormalities (ARIA) including brain edema and microhemorrhages (146). There trials have also often excluded those with comorbid cardiovascular disease, thereby limiting generalizability to most dementia cases among the elderly (129). Other therapies are under development including anti-tau agents, drugs that target glucose metabolism, or re-purposed drugs that aim to lower blood pressure (142).

Chapter 2 — Genetic Risk Prediction

In the previous chapter, I hinted that genetic factors play a role in these diseases through rare monogenic disorders (e.g., CADASIL in cSVD) or highly penetrant common variants like *APOE-ε4* in AD. Beyond these examples, growing evidence indicates a strong polygenic component underlying these complex conditions. Over the past few decades, advances in genetic epidemiology have begun to elucidate the genetic architecture and biological mechanisms, while offering new opportunities for risk prediction and stratification through PRS. This chapter provides an overview of genetic epidemiology and its contributions to cerebrovascular disease research, as well as the current evidence for PRS for stroke and dementia. Key terms in genetic epidemiology are found in **Box 1.**

Box 1: Key terms in genetic epidemiology

More detailed explanation of these terms and key concepts in epidemiology can be found in Chapter 2 of Genetic Epidemiology: Methods and Protocols (2018)

Allele: One of the variant forms of a gene or DNA base that can exist at a given genetic locus.

Single nucleotide polymorphism (SNP): The most frequent form of genetic variation, defined as a variation in a single DNA base. For a given SNP, the different possible bases that can occur at that position are called alleles.

Genotype: The combination of alleles an individual carries at a specific DNA site (e.g., AA, AT, or TT).

Linkage Disequilibrium: The residual correlation between specific alleles at SNPs on a chromosome that has not been broken down by historical recombination.

Monogenic inheritance: A pattern of familial recurrence for traits that follows the laws of Mendelian inheritance, typically involving rare genes with large effect sizes at the individual level. Diseases that follow this inheritance pattern are called Mendelian disease or monogenic.

Polygenic or complex disease: Diseases that have some genetic component but are not exclusively Mendelian. They are typically caused by a large number of genes with small to modest effect sizes and their complex interplay with environmental factors.

Heritability: The estimated genetic contribution to a phenotype, representing the proportion of variation in the trait that can be explained by genetic variation in the population.

Genome-wide association study (GWAS): Studies that test hundreds of thousands of genetic variants (like SNPs) across many genomes to find those statistically associated with a specific trait or disease.

2.1 Introduction to Genetic Risk Prediction

2.1.1 A brief history of genetic epidemiology

Genetic epidemiology is an interdisciplinary field making use of epidemiologic designs and statistical models to unravel the role of genetic determinants in health and disease and complex interplay with environmental factors. Today, genetic epidemiology contributes to precision public health and medicine, through providing genomic-based biological insights and characterizing individual genetic risk. I provide a brief history of the field that started with basic Mendelian genetics and family-based studies and now comprises large-scale GWAS with fine-grained characterization of the whole genome.

Box 2: Mendel's Laws

1. **The Law of Dominance:** This law explains that when two different alleles are present for a trait, one allele (dominant) determines the observable characteristic (phenotype) over the other allele (**recessive**).

2. **The Law of Segregation:** This states that during sexual reproduction (when an organism produces gametes), the two alleles of each parent randomly separate (segregate) so that each gamete receives only one allele.

3. **The Law of Independent Assortment:** This law was derived from dihybrid crosses and suggests that different pairs of alleles segregate independently of one another. This is generally true for genes located on different chromosomes, meaning they are unlinked.

Prior to sequencing the human genome, the field of genetic epidemiology was limited to familial studies of Mendelian genetics of rare traits and diseases. Mendelian genetics describes the patterns of trait or disease transmission that follow the laws established by Gregor Mendel, the founder of modern genetics in the mid-18th century, focusing on traits caused by variants in a single gene (**Box 2**) (147). Diseases following this clear pattern of **monogenic inheritance** are individually rare and manifest when the genetic factors exert a large effect on risk, often following clear dominant or recessive patterns within families (147). Due to their high penetrance, monogenic diseases have a major public health impact.

From the discovery of DNA's structure in 1953 to about 1980, genetic markers were limited to blood groups (ABO) or serum protein markers, and their precise locations on the human genome were unknown (148). Researchers performed basic linkage analyses in families with rare familial diseases to identify genes or chromosomal regions that contained disease mutations. This required the assumption that genetic markers segregated through families in a pattern similar to that of the putative disease gene, and that this gene was genetically linked (i.e. in close proximity) to the observed genetic markers (148). Indeed, the core concept of **linkage disequilibrium (LD)** was defined in this early period, describing that genetic variants in proximity to each other on the same chromosome are typically inherited together and spatially correlated within a population. In this

period we also saw early public health applications in genetics to monitor and control Mendelian diseases such as phenylketonuria (148).

Around 1980, the discovery of new polymorphic markers led to the development of the first genome-wide DNA marker panels. These panels consisted of approximately 500 simple tandem repeats—a sequence of two or more DNA bases that is repeated numerous times, immediately adjacent to one another—scattered throughout the genome that allowed for genome-wide linkage screening in multigenerational families with many affected relatives (148). This approach proved to be highly successful for monogenic diseases. For example, the discovery of amyloid precursor protein (APP) for early-onset familial AD led to a major breakthrough in our understanding of the disease at the time (149). To tackle the challenge of mapping complex or polygenic diseases, researchers developed non-parametric linkage models that utilize relatives, usually sibling pairs. In these models, affected relatives are examined to determine if they share genetic material more frequently than would be expected by chance (147,148). This approach identified breast cancer susceptibility genes (*BRCA1* and *BRCA2*) despite considerable loci heterogeneity within families, and led to successful public health screening programs that are still in effect today (150). During this period, public health programs incorporated knowledge and intervention strategies for Mendelian diseases, such as cystic fibrosis and Tay-Sachs disease, usually by performing genetic counselling in high-risk populations.

Nearing the end of the century, there was a technological shift toward the use of **single nucleotide polymorphisms (SNPs)** instead of short tandem repeats. Biallelic SNPs are the simplest form of human genetic variation, yet, by exploiting LD they can collectively they represent most unmeasured variation in a population (147). Effectively, SNPs “tag”, or represent, haplotype blocks which are regions of high LD bounded by recombination hotspots. By selecting certain "tagging" SNPs, one can infer the genotypes at other SNPs in the region that were not directly genotyped. This means the specific causal genetic variant does not need to be directly included on the panel for investigators to detect an association. At this time, candidate gene studies became popular, where researchers select “candidate” SNPs within a gene thought to have a mechanistic basis related to a phenotype.

The modern era of population genetics and genetic epidemiology really took off after the sequencing of the human genome was complete in 2001, providing a foundational reference for the field (151). At the same time, technological advancements in microarray genotyping chips, which now included hundreds of thousands of common SNPs, facilitated large-scale population genotyping. This progress enabled the first GWAS to explore the genetics of common complex

diseases (148). Essentially, GWAS uses basic regression models and statistical inference from epidemiology to systematically test the association of each individual genetic variant, usually SNPs, with a trait of interest across the genome.

The first GWAS, published in 2002, identified the lymphotoxin-α gene as a risk loci for myocardial infarction in a sample of 1,133 cases and 1,007 controls (152). In these early days however, very few loci for complex diseases were identified and their associations replicated (153). Further, GWAS only accounted for a small fraction of a trait's **heritability**, or total proportion of phenotypic variance that could be explained by genetic factors. Thus, it was recognized that there was a need for extremely large sample sizes to provide enough statistical power to identify associations with only modest effects on disease risk (e.g., odds ratios (OR) in the range of <1.5) (154). This has driven large collaborative efforts that combine multiple studies in meta-analysis of a common phenotype. Modern GWAS now include over a million people for multiple common diseases (155). Large-scale biobanks, such as the UK Biobank, have also been established to serve as comprehensive resources for studying the genetics of complex diseases.

Around the same time, next generation sequencing (NGS) technology was being used to sequence DNA at unprecedented speeds (156). Sequencing is performed on the level of the whole genome (whole genome sequencing, WGS), or the whole exome, including only protein-coding regions, (whole exome sequencing, WES). Leveraging these technologies, ongoing efforts were made to better characterize human genetic variation globally, and make standardized "reference panels" that can be used to enhance genetic information from SNP microarrays. Using low-coverage WGS, deep WES and SNP micro-arrays, the 1000 Genomes Project reconstructed the genomes of 2,504 individuals from 26 geographically and ethnically diverse populations around the world (157). They documented 84.7 million SNP in total, of which 20 million were considered common (minor allele frequency (MAF) > 0.5%) (157). This massive effort provides highly-accurate reference panels widely used for genotype imputation, a process used to greatly extend coverage across the genome by estimating genotypes at locations not directly measured by the genotyping array. Other more recent and comprehensive reference panels are provided by the haplotype reference consortium (HRC) (158) and the Trans-Omics for Precision Medicine Program (TOPMed) (159).

To date, just over 20 years after the first GWAS was published, GWAS have vastly increased our knowledge about the genetic architecture and biological mechanisms underlying many common, complex diseases (155). Beyond identifying risk loci, GWAS lays the foundation for further investigations into disease biology, allows for quantification of genetic architecture (e.g.

heritability estimation), enables risk prediction and drug development, and importantly drives the establishment of large-scale public resources for genomics research.

2.1.2 GWAS and heritability

2.1.2.1 GWAS

Large international GWAS efforts have uncovered numerous common variants associated with complex traits, including cerebrovascular diseases. In GWAS a case-control design is most commonly used to test if the number of allele copies (zero, one, or two) for a SNP is more frequent in disease cases than in unaffected controls. These SNPs may not necessarily be casually associated to the disease, but are likely in LD with the casual SNP and represent indirect associations. Conducting a GWAS involves a rigorous multi-step process that starts with quality control on the genotyped data, and handles potential biases in the form of population stratification and relatedness. Population stratification refers to systematic differences in genotypes between individuals of diverse ancestry (160). Over the course of history geographic isolation, interbreeding, and adaptation caused human populations to differentiate from each other. This caused divergent random genetic drift in geographically isolated populations, causing allele frequencies to randomly change overtime (160). Relatedness between individuals in a population also naturally introduces bias in GWAS by violating the statistical assumption that all individuals are independent. It can be either known based on family structure or unknown (i.e. cryptic relatedness). Handling of these biases are described in more detail in **Appendix 1.** Multiple testing burden is also important to consider, with the standard being the Bonferroni threshold of $p < 5x10^{-8}$, referred to as the genome-wide significance level (GWS). This threshold is based on the finding that there are approximately 1 million independent common genetic variants on average (155). The Bonferroni threshold is derived by dividing the standard significance threshold by the number of independent tests performed ($0.05/10^6=5x10^{-8}$). GWAS results are usually visualized by a Manhattan plot (**Figure 11**), where multiple SNPs in LD on the same locus represent distinct peaks.

It is also necessary to attempt replication of GWAS signals in independent cohorts (161). Therefore, GWAS are most often divided into a 'discovery' phase where GWS SNPs are identified in a sufficiently large sample, and 'replication' phase. The replication GWAS, usually smaller than the discovery GWAS, is used to validate the initial associations at more liberal significance thresholds (e.g. $p<0.05$) and refine effect estimates. Importantly, the replication GWAS must be entirely independent and match the ancestry of the discovery GWAS. In addition to using a the strict GWS multiple testing threshold, the use of an independent replication GWAS is the gold

standard for accounting for false discovery and the 'winner's curse', the phenomenon where effect sizes tend to be overestimated in the initial discovery GWAS (155).

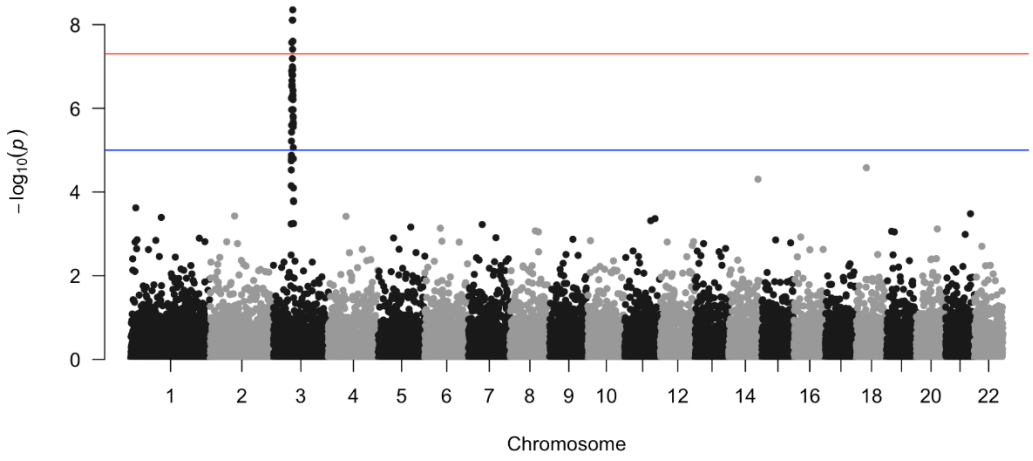


Figure 11. Example of a Manhattan plot from GWAS. It represents the association's statistical significance as $-\log_{10}$ (*P*-value) in the *y*-axis against chromosome coordinates in the *x*-axis. Each dot corresponds to one variant, and the higher a variant (−log10 of the association *P*-value) the more significantly associated it is with the phenotype. The red line represents the GWS level ($p=5x10^{-8}$) and the blue line is a suggestive threshold (e.g. $p=1x10^{-6}$).

The primary output of GWAS—summary statistics that contain a list of SNPs and their effect sizes and associations with a given trait—are routinely shared in public databases such as the GWAS catalog (https://www.ebi.ac.uk/gwas/) which has over 146,000 summary statistics across 7,426 publications (162). Publicly available GWAS summary statistics foster reproducibility and are leveraged to conduct a suite of post-GWAS investigations across a range of applications. Beyond SNP-based heritability estimations and PRS, other post-GWAS methods central to modern genetic epidemiology are described in **Appendix 2**.

2.1.2.2 Heritability

The concept of heritability is core to genetic risk prediction. Heritability is defined as the proportion of variation in a phenotype that can be explained by genetic factors this is also called "broad-sense" heritability. Heritability can be parsed into additive effects, or non-additive effects which include dominant effects (i.e., gene-gene interactions within a locus) and epistatic effects (i.e., gene-gene interactions between loci). The variance explained only by additive genetic effects is called "narrow-sense" heritability which represents the degree that genes transmitted by parents determine the traits of their children (163).

Traditionally, broad-sense heritability is estimated from twin studies by comparing genetically identical twins to non-identical twins who like most siblings, share approximately half of their genome. Family-based methods were also developed to overcome possible environmental bias in twin studies and incorporate all family information to produce more precise estimates (163). These

studies established an initial benchmark for the genetic contribution to many traits (164–166). Importantly, narrow-sense heritability is what could be theoretically captured by GWAS. A key insight in genetic epidemiology was that one can estimate a portion of narrow-sense heritability by utilizing all genotyped SNPs and exploiting subtle variation in genetic relatedness among unrelated individuals (167). More specifically, this heritability is called SNP-based heritability (h^2_{SNP}) which is defined as the proportion of phenotypic variance explained by all the genotyped or imputed SNPs, and is considered to be lower bound of narrow-sense heritability (167). There are methods available to measure h^2_{SNP} from individual-level genotyped data or using GWAS summary statistics (168,169). These methods depend on distinct model assumptions that may lead to differences in estimation (168). An important extension of these methods is the ability to estimate the genetic correlation or the shared genetic influences between two traits (170).

There were early hopes that with large-scale genotyping and GWAS we would be able to capture a substantial portion of narrow-sense heritability. However, significant common variants identified in GWAS can only explain a small fraction of heritability for most complex traits, leading to what is known as the 'problem of missing heritability' (166). This problem has persists even if all genotyped SNPs are used beyond only those associated in GWAS (171). For example, the narrow-sense heritability of height and schizophrenia are both estimated to be 70-80%, but h^2_{SNP} is only about 50% and 20% respectively (167,171,172). This issue has been widely discussed and many propositions have been made about where to "find" the missing heritability. Some possible sources include (164,173):

- **Missing variants** that are not captured by genotyping arrays. These include non-SNP variants such as structural variation, rare large variants exceeding 500 kilobases, copy-number polymorphic duplications, or rare variants collectively that may be missed by GWAS genotyping arrays. Beyond GWAS, WGS can help to address missing heritability by capturing rare and structural variants that comprehensively represent a trait's genetic architecture (174).
- **Additional common variants with small effects:** Complex traits are highly polygenic with many genetic variants contributing weak individual effects, and GWAS are not sufficiently powerful to detect these effects.
- **Gene-gene interactions** or non-additive effects are not modeled or captured. This is the concept of epistasis where the effect of one gene is modified by the effect of another, and that there are 'rare' combinations of common variants that contribute to disease.

- **Gene-environment (GxE) interactions** are not modeled or captured. Lifestyle changes and environmental exposures may alter the distribution of genetic effects in some people. Capturing these interactions would require identifying and modifying the relevant environmental factors, and sufficient statistical power to detect interactions.
- **Overestimation of heritability in benchmark studies.** Twin studies assume that twins share the same environment. This is likely not the case in practice where hidden environmental structure may interact with genes. To this effect, independent family-or pedigree-based studies and methods robust to environmental bias typically produce lower heritability estimates (163).

2.1.3 Polygenic risk scores

PRS, by leveraging GWAS data, quantify an individual's genetic liability for a trait or disease across the whole genome and are thus the principal tool for genetic risk prediction. While h^2_{SNP} estimation forms the basis of quantifying genetic contribution to disease, it does not provide individual-level risk estimates that can be used for prediction, and this is where PRS are valulable. Shortly after the first GWAS were conducted, an approach was proposed to assess the genetic risk of a disease in healthy individuals based on the genome-wide SNP panels used in GWAS (175). Notably, the authors proposed that individual risk alleles underlying complex disease alone have small effect sizes, but when considered as a cumulative effect could allow the identification of individuals at high risk (175). Their simulation studies estimated that, depending on the prevalence and heritability of disease, those in the top 5% of predicted genetic risk were three to seven times more likely to develop the disease (175). This study provided the first theoretical evidence that common genetic variants could be aggregated to identify individuals at highest genetic risk, and could be a better predictor of genetic predisposition than family history. With this theoretical basis as a foundation, the first successful polygenic risk prediction was applied to schizophrenia in 2009, showing a polygenic score derived from schizophrenia GWAS was associated to the disease in a replication sample, despite only a few of the individual SNPs being statistically significant (176). This study established important evidence that common complex disease were indeed in part caused by the additive, small effects of common risk alleles in the population and justified larger GWAS sample sizes (154). These scores are now commonly referred to as PRS and represent the principal genetic risk prediction tool derived from GWAS.

Formally, PRS combine risk alleles of common, independent SNPs, usually weighted by effect sizes from GWAS, into one value to represent an individual's genetic liability to a trait. PRS can be generated in independent target cohorts with genotyped data, and used for association testing

and risk prediction. A full description of PRS methodology is provided in **Chapter 3 (Methods)**. Importantly, the theoretical upper limit of PRS predictive ability is in fact h^2_{SNP}, in other words, if the effects of SNPs were estimated in a GWAS without error, then PRS could predict the phenotype of individuals in the target data with variance explained equal to h^2_{SNP} (177,178). In reality, the predictive power of PRS is typically much lower than h^2_{SNP} due to estimation error in effect sizes and differences between the samples included in a GWAS and the target cohort (177). As GWAS sample sizes increase, the variance explained by PRS will also increase and approach the h^2_{SNP}.

In earlier studies, PRS was primarily useful for association testing to investigate whether a genetic signal was present for a trait among an ensemble of selected markers (154). Testing for association between the PRS and other traits of interest in an independent sample can also reveal a common genetic basis for related traits. The most enticing use of PRS is for individual risk prediction as PRS have the advantage of being measurable from the time of birth to provide the earliest possible indication of disease risk. However, it was argued that much larger samples, to the magnitude of hundreds of thousands of individuals, were required to accurately estimate genetic effects and provide useful levels of prediction (154). The promise of PRS to provide clinically meaningful risk prediction has in part motivated the vast efforts to increase sample size of GWAS for many common diseases.

2.2 Genetic Architecture of Cerebrovascular Diseases

2.2.1 Monogenic disorders

Most monogenic disorders are autosomal dominant or autosomal recessive. Autosomal dominant inheritance occurs when a mutation in only one copy of a single gene is sufficient to cause the disease. Autosomal recessive inheritance requires mutations in both copies of the gene to cause the disorder.

In the early 1990's, family-based studies identified three highly penetrant mutations in *APP*, *PSEN1*, and *PSEN2* that cause early-onset, autosomal dominant AD (179). Mutations in *APP* account for approximately 14% of early-onset autosomal dominant cases of AD (180). Mutations in *PSEN1* are the most common cause of early-onset familial AD, accounting for about 80% of cases, while *PSEN2* mutations account for about 5% (180). These mutations often lead to altered Aβ production or a shift in Aβ ratios, supporting the hypothesis that Aβ accumulation is central to AD pathogenesis.

The most common and well-studied monogenic disorders of stroke are related to cSVD (181). CADASIL is by far most common monogenic form of cSVD resulting from mutations in the *NOTCH3* gene (181). In CADASIL patients, WMH are usually present from age 30, and affected individuals usually have early-onset of migraines, and develop lacunar stroke, typically after age 40, with increasing cognitive impairment as they age (181). The autosomal recessive form (CARASIL) is substantially more rare and is due to mutations in the *HTRA1* gene with similar clinical consequences to CADASIL (181,182). More recently, it has been found that only a single *HTRA1* mutation can also lead to cSVD in an autosomal dominant pattern (CADASIL2) (181). Mutation in *COL4A1/2* which encodes a type IV collagen alpha protein, can present with autosomal dominant cSVD as well. Type IV collagen is an integral component of the basement membrane of blood vessels, and dysfunction in its formation predominantly predisposes to ICH, but also lacunar stroke, with mean age of onset in the 30's (181,183). Other extremely rare forms of monogenic cSVD exist including mutations in the *CTSA* or *TREX1* genes (181,184,185). Lastly, hereditary CAA predisposes to high risk of ICH caused by mutation mostly in *APP*, and less frequently mutation in *CST3, BRI, ITM2B, GSN*, and *PRNP* (186).

2.2.2 Common genetic variation in cSVD

GWAS meta-analyses have been conducted across several cSVD MRI-markers (e.g., WMH, PVS, silent brain infarcts, and CMB) discovering more than 50 risk loci associated with these traits (13–16,187). Additional loci have been found to be associated with novel dMRI markers of cSVD, like fractional anisotropy, mean diffusivity, and PSMD (16). Interestingly, common variants in both

HTRA1 and *COL4A1/2* are associated with the sporadic forms of cSVD – indicating a continuum with monogenic disease (181). There is also consistency in associations of genetic loci across MRI markers of cSVD with 34% showing associations with at least two distinct markers (16). These loci point to several key biological pathways underlying cSVD pathology, most notably involving extracellular matrix structure and function (16). Other risk loci are involved in myelination, membrane transport and blood-brain barrier dysfunction (16). Additionally, roughly half of the cSVD loci are shared with established vascular risk factors (e.g., blood pressure traits, T2D, LDL cholesterol) suggesting both shared mechanisms involved in overall vascular health, but also independent cSVD mechanisms that aren't mediated by these risk factors (16). More recently, GWAS on NODDI metrics, revealed several loci (e.g. *VCAN* and *STAT3*) that were implicated in young adults (<35 years), providing genetic evidence that cSVD takes root early in life (71).

Family-based estimation methods for cSVD-related WMH have indicated a heritability of 55-72% (188–190). For both cSVD MRI-markers and stroke, h^2_{SNP} estimations using individual-level genotyped data for all available SNPs (with GCTA software (191)) have yielded high heritability estimates in line with family based studies. Across cSVD MRI-markers h^2_{SNP} was 59% for PVS, 55% for WMH, and 48% for lacunar brain infarcts (192).

2.2.3 Common genetic variation in stroke

Stroke GWAS efforts have been ongoing since 2007, and have been primarily focused on ischemic compared to hemorrhagic stroke. The latest GWAS meta-analysis for ischemic stroke, conducted by the GIGASTROKE consortium, identified 89 independent loci associated with IS, and was comprised of 110,182 stroke cases (across five ancestries, 33% non-European) and 1,503,898 stroke-free controls (10). The same study also looked at ischemic stroke sub-types individually, identifying 15 loci associated with LAS, 13 with CES, and 10 with SVS. Several of the loci identified are highly implicated in coagulation and thrombosis pathways, and others agree with findings from GWAS of cSVD MRI markers with loci involved in the extracellular matrix and BBB integrity (10). A substantial portion of stroke risk loci overlap with traditional vascular risk factors (64%), particularly, 33 loci which were also associated with blood pressure traits (10). Comparatively, the largest GWAS for ICH to date only included 1,545 ICH cases and 1,481 controls (11). Only two common genetic loci have been identified for ICH, most notably *APOE* where both the ε2 and ε4 alleles increase lobar ICH risk with an odds ratio of 1.8 and 2.2, respectively, versus a reference group (*APOE ε3/ε3*) (11,186).

The h^2_{SNP} of stroke was between 16-40% for ischemic stroke depending on the sub-type (193), and estimated at 44% for ICH (194).

2.2.4 Common variation in dementia

The *APOE* gene, identified through linkage studies in the 1990's, is the most ubiquitous genetic risk factor for AD-type dementia. *APOE* occurs as three possible isoforms: *APOE-ε2*, *APOE-ε3*, and *APOE-ε4* (see **Table 3** for prevalence of each). Compared to *APOE-e3/e3* genotypes, carrying one copy of ε4 increases the risk approximately 3-fold (for heterozygous carriers), while carrying two copies increases the risk approximately 8- to 10-fold (for homozygous carriers) (179).

APOE Genotype	Global prevalence (approximations)
APOE-e3/e3	54%
APOE-e3/e4	25%
APOE-e2/e3	12%
APOE-e2/e2	1%
APOE-e2/e4	4%
APOE-e4/e4	4%

Table 3. *APOE* global prevalence from Andrews et al. (2023) (195)

Meanwhile, the ε2 allele interestingly conveys a decreased risk for late-onset AD and delays age at onset (179). There are clear mechanisms that link *APOE* with AD as *APOE* strongly affects Aβ clearance and aggregation in the brain (179). In the first GWAS for AD of modest samples sizes (N <2,000), *APOE* was the only locus to consistently reach genome-wide significance (196). To date AD-GWAS, comprising 111,326 AD cases and 677,663 controls in the latest meta-analysis, have identified 75 independent risk loci beyond *APOE* (12,195). AD is unique in that it includes an extremely, high penetrant common variant in *APOE-ε4* among a background of polygenic influences. The vast majority of these variants are common alleles with individually small effect sizes that collectively contribute to the genetic liability for late-onset sporadic AD. *APOE* is primarily implicated in Aβ clearance and aggregation, cholesterol metabolism, and microglial-mediated clearance of cellular debris (efferocytosis) (179,195). Other AD risk loci are involved in neuroinflammation and innate immune response, specifically in processes involving myeloid cells (e.g. microglia), and also in efferocytosis and cholesterol metabolism (195). Recently, other GWAS have importantly investigated all-cause dementia (ACD) and VaD involving 46,902 ACD and 8,702 VaD cases, respectively, in a cross-ancestry meta-analysis (197). They replicated 10 known AD loci, reaffirming that AD pathology likely drives the majority of common genetic risk for ACD, but also identified three novel loci for ACD (197). *APOE* was the only variant to reach GWS for VaD, highlighting its importance beyond Aβ aggregation through implication in BBB and microglial function (197), which is also supported by associations with cSVD MRI-markers,

WMH and CMB (198,199). There is also genetic evidence that VaD is linked to blood pressure, and WMH (197). Overall, the genetics of VaD is much less established than AD likely due to its multifactorial nature and heterogenous definition with respect to clinical and pathological criteria. Rather, the genetics of VCI is understood mostly in the context of known pathological risk factors like cSVD, CAA, and stroke (200).

Heritability of late-onset AD is estimated to be approximately 60-80% in twin studies (201,202), and there are no available estimates for VaD or other dementia types. The missing heritability problem seems more apparent for AD with h^2_{SNP} estimates ranging from 13-31% (203). Further, the common *APOE-ε4* allele may account for up to a quarter of total heritability (203,204). Estimates of h^2_{SNP} for VaD are more limited with one study offering an imprecise estimate of 6.1% [95% CI 3.2, 21%] (197).

2.2.5 A note on SNP-based heritability estimations

It is essential to note that the above h^2_{SNP} estimations are based on individual-level genotyped data, and it is more and more popular to measure h^2_{SNP} based on GWAS summary statistics, such as LD score regression (LDSC) (169). However, this method yields substantially lower heritability estimates than GCTA for these diseases. Based on the most recent GWAS summary statistics for stroke and cSVD-markers, h^2_{SNP} estimations are much lower: ischemic stroke sub-types have been estimated to be very low 0.2-0.5% (205), and 12.5% for WMH and 6.5% for PVS. A similar disparity is noted for AD, to the extent that using both methods on the same exact dataset result in estimates of 17% for LDSC and 31% for GCTA (206). While both methods assume a uniform heritability model, each SNP contributes an equal portion of heritability, they use different core statistical procedures involving relatedness matrices (GCTA) or LD (LDSC) (207). At the fundamental level, the uniform heritability model is likely overly simplistic, and for instance, doesn't consider complex relationships or patterns between SNPs in LD or the impact of confounding (207). Further, summary statistics-based methods are reliant on the quality of underlying data in which confounding was adequately controlled for in the primary GWAS. Though the GCTA or individual genotyped-based estimates above tend to align more with family-based estimates, this method is susceptible to confounding inflation by poorly accounting for underlying population structure, and also tends to overestimate heritability in high-LD regions (168). The variable, and sometimes quite low, estimated h^2_{SNP} from GWAS summary statistics-based methods likely provides a closer estimate to the limits of predictive ability for a PRS derived from a given GWAS.

2.3 Polygenic Risk Scores Applied to Cerebrovascular Disease

There have been great efforts in recent years to optimize and demonstrate the clinical utility of PRS across many common diseases. A landmark paper in 2018, using data from the UK Biobank, provided strong evidence that extreme PRS values conveys risk comparable to rare monogenic mutations and adds accuracy to clinical risk predictors for common diseases (208). Namely, for coronary artery disease (CAD), participants in the top 8% of the PRS score distribution had a threefold higher risk for CAD relative to the rest of the population. This 8% prevalence is 20-times higher than the carrier frequency (0.4%) of a familial hypercholesterolemia mutation, which also conveys a three-fold increased risk for CAD (208,209). It has also been demonstrated that PRS are largely capturing risk independent of traditional risk factors (17).

A recent review summarized the evidence for the potential of PRS for the prevention of several common cardiometabolic diseases including CAD, AF, type 2 diabetes mellitus (T2D), and stroke (22). The review highlights that PRS for CAD, AF, and T2D show consistent evidence for improved prediction when integrated into existing clinical risk scores. Stroke PRS, however, lags behind PRS for other cardiovascular diseases in terms of predictive ability (**Figure 12**) (22).

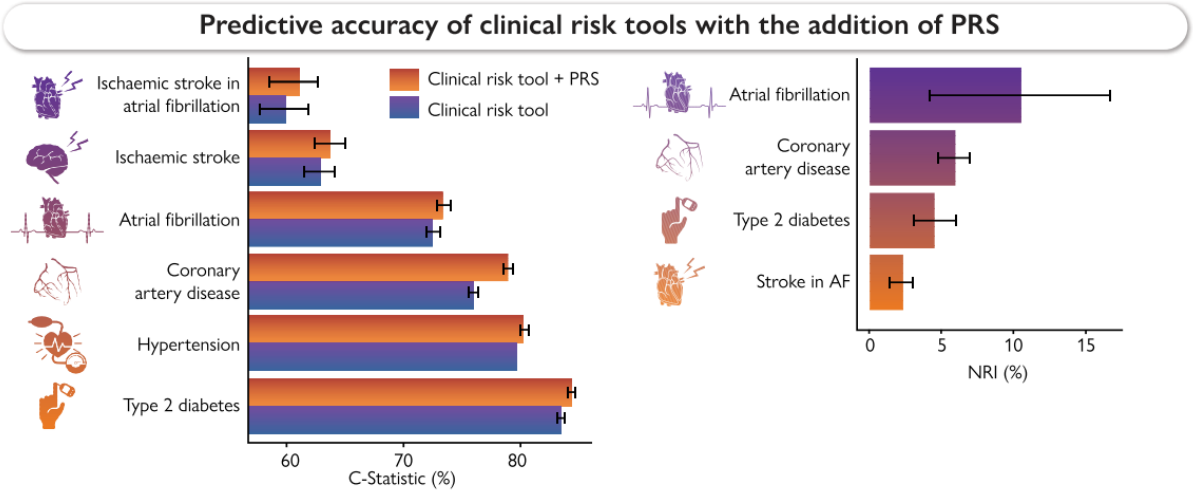


Figure 12. *From O'Sullivan et al. (2023).* This review summarized the current evidence for PRS across cardiometabolic diseases. A C-statistic of > 70% is usually a benchmark for adequate predictive performance. Stroke PRS clearly lags behind others, including in patients with AF where stroke PRS could be particularly useful to guide anticoagulation.

PRS for cerebrovascular disease overall are comparatively less advanced. Additionally, most existing PRS evidence is for stroke and dementia rather than cSVD, since these later-life conditions have comparatively larger GWASs available.

2.3.1 PRS for cSVD MRI-markers

Among MRI markers of cSVD, WMH are the only trait with GWAS data sufficiently powered to construct PRS. Existing WMH-PRS—typically based solely on genome-wide significant variants—have been used only sparingly in association analyses. Prior studies have reported associations between WMH-PRS and dMRI measures, such as PSMD and NODDI metrics, in young adults (13,71). However, current WMH-PRS have not been optimized for prediction, either for WMH burden itself or for downstream outcomes such as stroke and dementia.

2.3.2 Stroke PRS

Early PRS models for ischemic stroke incorporating smaller numbers of SNPs (e.g. 90-SNP PRS) (221), were shown to be associated with stroke independently of clinical and lifestyle factors, but improving prediction beyond these factors was limited (21,22). Stroke prediction faces particular challenges due to its heterogeneity. To address this, one can further expand coverage to include many SNPs beyond genome-wide significance, and additionally attempt to capture independent genetic signals associated with stroke risk through intermediate phenotypes like vascular risk factors in a "meta-GRS" or integrative-PRS (iPRS) framework (211). A 3.2 million SNP ischemic stroke iPRS (IS-iPRS), derived from PRS from 14 vascular risk factors and stroke-subtype GWAS, doubled the strength of association with ischemic stroke compared to the 90-SNP PRS, and captured risk not quantified by established risk factors with predictive ability equivalent to SBP (211). The most recent IS-iPRS, as part of the GIGASTROKE project, further integrated both ancestry-specific and cross-ancestry PRS for stroke sub-types in addition to vascular risk factors, and further improved stroke prediction in the general population and across clinical trials over the previous IS-iPRS model (10). The authors also derived an iPRS for East Asians, and demonstrate that iPRS models derived from cross-ancestry stroke GWASs had a higher predictive ability compared with iPRS models derived from ancestry-specific stroke GWASs both in Europeans and East Asians (10). The IS-iPRS for Europeans showed significant albeit weaker association with IS incidence in African-American or indigenous African ancestry (10).

Since the GWAS data is severely lacking in power for ICH, there has been limited studies investigating PRS. However, a similar integrative PRS approach, incorporating PRS for 21 ICH-related traits across 2.6 million SNPs was consistently associated to ICH risk in the general population, and was able to identify high-risk individuals in the extremes: the top 1% conveyed 5-fold higher odds of ICH compared to the rest of the sample. Similar to IS, they also showed ICH-iPRS was independent of, and improved prediction beyond clinical risk factors (212). Another study also showed that a single-trait ICH-PRS showed potential benefit to identify those at risk of

ICH among anticoagulant users on vitamin-K agonists (warfarin) (213). *APOE* genotype was also found to modulate bleeding risk among users of DOACs (214). In fact, in stroke prevention, PRS could be especially useful to help guide clinical decision-making when the indication or duration for anticoagulation therapy is controversial or uncertain. Along these lines, a simple 32-SNP IS-PRS demonstrated ability to identify individuals with AF who are at increased stroke risk despite having moderate clinical risk scores (e.g., CHA_2DS_2-VASc scores of 2) (215).

2.3.3 Dementia PRS

APOE accounts for a large portion of the genetic contribution to AD risk, yet there is evidence that AD is highly polygenic and AD-PRS have shown ability to improve risk prediction beyond *APOE* (216,217). One study, estimating heritability in twins to be 71%, modeled genetic risk with *APOE* and an AD-PRS (excl. *APOE)* showing they capture 9.3% and 2.1% of variability in AD among twins (201). Given the outsized contribution of *APOE*, studies have shown it is optimal to model *APOE* separately from all other common risk alleles with small effects (217,218). Indeed, one study indicated the best prediction accuracy (area under the curve, AUC=74.1%, R^2 =0.24) was consistently achieved by a model combining an AD-PRS (M=68,516 SNPs) with directly genotyped APOE ε2 and ε4 alleles as separate predictors (217). AD-PRS also captures risk within *APOE* genotype groups, especially considering the extremes of PRS where the highest 2% had a significantly higher risk of AD within strata of each *APOE* genotype group (OR>2) (219). PRS may also be effective at identifying distinguishing risk and earlier age-of-onset among *APOE-ε4* carriers (219,220), and has been further associated with AD in individuals without *APOE-ε4* (221). There is also evidence of age-dependent effects of *APOE* versus AD-PRS, where the effect of *AD-PRS* is stronger at older ages as the ε4 allele decreases in frequency (217). However, more investigations are needed to determine if there is an additive or non-additive effect for *APOE* and AD-PRS, and how genetic risk should be comprehensively modeled for AD.

Unlike for stroke, a metaGRS or iPRS approach has been applied for AD-PRS, but did not show any improvement over a simple, single-trait AD-PRS (222). An alternative PRS methodology that has been applied to AD is to restrict PRS to gene sets representing distinct biological pathways that may play an important role in identifying mechanisms of disease development. Pathway-specific PRS are not generally more predictive than genome-wide PRS or *APOE*, they show independent associations with both AD risk and amyloid deposition (223). A more recent pathway-based investigation revealed pathway-specific PRS showed unique association with structural changes across brain regions based on neuroimaging (e.g. cortical thickness) (224). This approach

has potential to disentangle the genetic contribution to underlying different biological processes that define AD, and also potentially identify novel implicated pathways.

Most studies have also only considered AD-PRS in relation with an AD outcome rather than ACD or VaD. One study showed AD-PRS predicts ACD and VaD (221,225), though this association is attenuated compared to AD-only (OR_{AD}=1.70; OR_{ACD}=1.36; OR_{VaD}=1.21). Two other studies showed association with ACD in healthy, older adults (>70 years) (220,221). The consistent, attenuated association of AD-PRS with ACD and VaD is likely due to patients often exhibit underlying mixed dementia that is not captured in current dementia classifications. Some investigations have also been made into the association of AD-PRS and cognitive decline, and show evidence of AD-PRS to predict progression to dementia among those with MCI (226), but not in healthy older adults (220). These findings indicate that AD-PRS may be particularly useful in higher risk populations, for example, memory clinics patients that could be preferentially selected for clinical trials.

2.3.4 Potential utility of PRS

From a primary prevention perspective, PRS captures a component of lifetime risk and can identify high-risk individuals prior to the emergence of conventional clinical risk factors (17). A PRS-based risk stratification strategy could be used for early identification of individuals at high risk for cardiometabolic disease (**Figure 13**) (227). It has been also demonstrated that integrating CAD-PRS into clinical risk stratification tools (e.g QRISK-2, ASCVD-PCE, and SCORE2) has added benefit to identify high-risk individuals across multiple ethnic groups (22,228,229), and may be specifically more powerful in younger individuals (230). This would enable healthcare providers to offer preventative therapy, such as statins, to individuals not identified via current risk assessment guidelines. Further, CAD-PRS has shown the ability to identify those most likely to benefit from treatment. A post-hoc analysis of statin prevention trials used a 27-SNP PRS (231) to risk stratify participants based on genetic risk, and individuals in the highest quintile of genetic risk benefited most from statin therapy compared to those in the lowest quintile (231). Additionally, those with high PRS for CAD receive greater absolute and relative risk reduction from PCSK9 inhibitors compared to those with low PRS or high clinical risk alone (232,233). Likewise, for T2D, individuals with higher PRS show greater benefit from lifestyle interventions (234), and could be additionally targeted for regular HbA1c screening (22). Overall, there is less evidence for secondary prevention use cases of PRS for cardiometabolic disease where PRS could help personalize pharmacological treatment or anticipate disease complications. Potential clinical applications of PRS in primary and secondary prevention of cardiometabolic disease are summarized in **Table 4** (22).

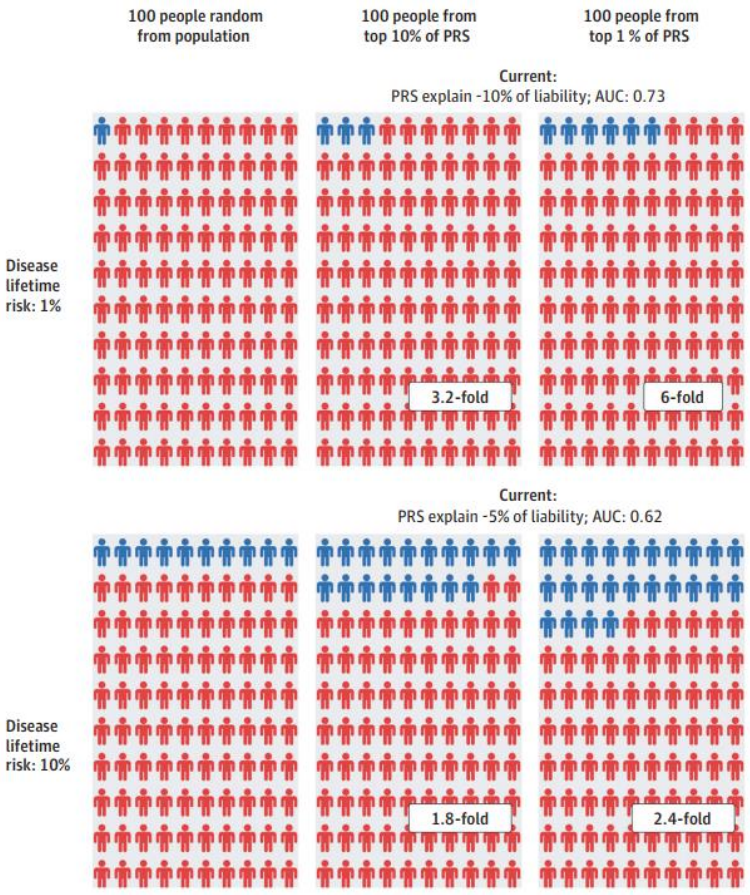


Figure 13. Potential of PRS to be used for population stratification. Selecting people from the top 10 or 1% of a distribution of a PRS even if it accounts for <10% of variance (or liability) to disease will help identify a non-negligible number of people. PRSs with lower disease lifetime risk in the population require stronger PRS to provide meaningful prediction. *Adapted from Wray et al. (2021)*

Primary prevention	Promote a healthy lifestyle	Communicating a patient's risk may promote a healthier diet and increased exercise, although data examining this are mixed.
	Enhanced prediction	For a number of cardiometabolic diseases, the addition of PRS enhances the prediction of classification.
	Earlier therapeutic interventions	Patients at high risk may benefit from early interventions, such as those at high risk of CAD.
	Robust risk factor control	Many cardiometabolic diseases are risk factors for other cardiometabolic diseases, such as hypertension for coronary artery disease. Identification of those at high risk could facilitate more robust risk factor control, e.g. lower blood pressure target for those at high risk of CAD.
	Personalized screening	Those at higher PRS may benefit from earlier routine screening to facilitate earlier intervention, such as regular HbA1c screening. Although long-term data is required.
	Cascade screening	For patients who have a PRS that is similar to that inferred by monogenic risk. This may mean families may benefit from screening, as is recommended for monogenic diseases
Secondary prevention	Drug response	For some cardiometabolic diseases, there is an array of medication options. PRS may indicate which medication or which combination of medicines is most beneficial.
	Disease progression	At diagnosis, the trajectory of disease is largely unknown, PRS could indicate which participants will develop mild or severe disease.
	Disease complications	Prediction of complications from cardiometabolic conditions, such as ischaemic stroke in patients with atrial fibrillation.

Table 4. Proposed clinical utility of PRS in cardiometabolic disease. *From O'Sullivan et al. (2023).*

2.3.4.1 PRS for drug development

Genomics, in general, show great potential to accelerate drug development and discovery through prioritizing targets identified in GWAS and establishing evidence for a casual links between drug target and a disease (235). PRS could also play a substantial role in enhancing the efficiency of clinical trials through prognostic (i.e. increased risk of disease) and predictive enrichment (i.e., increased benefit from intervention) (235,236). With prognostic enrichment, individuals with higher genetic risk of disease are likely to experience earlier disease onset, and by enrolling these individuals trial duration and required sample sizes may be reduced (236). Likewise, for predictive enrichment, a PRS can also identify individuals with increased probability to benefit from intervention (i.e., optimal drug responders), which would reduce sample size needed to demonstrate a positive effect (236). Predictive enrichment could be important for proof-of-concept trials whereby if a drug fails to show efficacy in a selected population, it has a very low chance of being approved (233). Again, for statin therapy, PRS demonstrated both prognostic enrichment—the rates of developing heart disease in the placebo group was 19.6% for those in the top quintile of the score versus 12.9% in all others—and predictive enrichment, where a 44% relative risk reduction was estimated in those with high PRS compared to only 24% in the remainder of the participants (236). Lastly, PRS can be useful for identifying rare variants that could lead to drug discovery (235). For instance, individuals who present with a disease but have a very low PRS may be harboring a rare, disease-causing variant representing a novel target. Reciprocally, individuals with very high PRS who do not develop the disease may have rare disease-protective variants that can guide the development of drug inhibitors to mimic loss-of-function variation.

2.3.5 Caveats and challenges to PRS implementation

PRS face a range of scientific, clinical, and societal challenges that limit their broader clinical utility for complex diseases. The most widely recognized barrier is their limited generalizability across ancestries, driven largely by the disproportionate representation of individuals of European descent in GWAS (~80% of all participants) (237). Beyond the need to expand GWAS sample sizes and improve the precision of effect-size estimates, PRS are also affected by sample- and method-dependent variability, leading to a lack of standardization and inconsistencies among PRS for the same trait (238,239). In addition, there are few established clinical guidelines to help healthcare professionals use PRS in decision-making (240). These challenges—and their implications for PRS in cerebrovascular diseases and more broadly—will be explored in greater detail in Chapter 7.

Rationale and Aims of Thesis

In 2025 people are living longer than ever before, leading to a growing burden of non-communicable diseases – particularly age-related conditions such as stroke and dementia. The global burden of these conditions is projected to increase two- to three-fold by 2050, disproportionately affecting low- and middle-income countries (1,2). Vascular risk factors—including hypertension, physical inactivity, obesity, smoking, and diabetes—strongly influence the risk of both stroke and dementia, many of which are also projected to increase in prevalence. Underlying both conditions is cSVD**,** a common yet underrecognized pathology characterized by changes in the structure and function of small brain vessels. cSVD is a known cause of stroke and is the primary pathological substrate of vascular contributions to cognitive decline and dementia (7,8). As such, targeted prevention of cSVD and the management of associated vascular risk factors—especially hypertension—represent major opportunities to prevent stroke and dementia in the general population (9).

Fortunately, in the era of precision medicine and public health, there is increasing recognition of the complex interplay between genetics, lifestyle, and the environment. Cost-effective genomic technology has made germline genome profiling more accessible and generated large collections of genomic data that can be leveraged to develop precision prevention strategies. GWAS exploit these rich datasets to find associations between common genetic variants and a number of traits, diseases, and their risk factors. To date, large-scale GWASs have identified ~90 loci associated to stroke (10,11), over 75 loci for dementia (12) and over 50 loci across MRI-markers of cSVD (13–16,187). Using data generated from GWAS, we can then quantify genetic predisposition to these conditions and associated traits through combining small variant effects in PRS. PRS are emerging as powerful genomic tools that have potential to guide disease prevention and management by stratifying disease risk and predicting disease prognosis and response to therapy (17,18). Integrated risk tools that combine PRS with clinical risk scores are currently being evaluated for clinical use. PRS for coronary heart disease have shown promise when added to the QRISK-2 clinical model (229), and adoption of such tools have been found to be both feasible and acceptable among providers and patients (241).

In the context of dementia, PRS has shown to predict disease risk—particularly for AD—beyond the *APOE-ε4* allele (19,20,217), its strongest genetic risk factor. However, these PRS require further validation in longitudinal settings across diverse populations and in high-risk groups. Additionally, incorporating the vascular contribution to dementia may improve the predictive

performance of PRS. For stroke, PRS are robustly associated with ischemic stroke in population and trial data (10,21,211), but their added value beyond clinical risk models remains unclear. To date, PRS has primarily been applied to primary stroke prevention contexts, showing particular utility to guide statin therapy (242,243). However, its role in secondary prevention—such as predicting treatment response or disease progression in high-risk patients—is underexplored (22). In ICH survivors with AF, the PRESTIGE-AF trial showed direct oral anticoagulants effectively reduced ischemic stroke risk but this benefit was offset by increased risk of major hemorrhagic complications (102). PRS could help balance ischemic and hemorrhagic risk, guiding anticoagulation therapy in these highly vulnerable individuals. Lastly, the genetic architecture of cSVD has also been explored using MRI-markers—most extensively for WMH (13). PRS derived from WMH could enhance early risk prediction of stroke and dementia and provide insight into underlying pathways relevant to early mechanisms and later-life complications of cSVD.

In light of the current evidence, we lack clarity on which populations and clinical contexts would most benefit from PRS applications in cerebrovascular disease. There is also an opportunity to refine and develop PRS models that more comprehensively capture polygenic risk of these complex and heterogenous diseases. This thesis aims to address these gaps by evaluating and developing PRS in the context of stroke, dementia, and cSVD. Specifically, we aim to:

1 **Develop an integrative PRS for dementia**, incorporating both neurodegenerative and vascular traits to improve risk prediction and stratification in older community-persons, memory-clinic patients, and in a multi-ancestry biobank.

2 **Apply existing PRS models** relevant to stroke in a secondary prevention context, using data from the PRESTIGE-AF trial of ICH survivors with AF to evaluate the ability of PRS to predict ischemic and recurrent ICH risk and identify those who benefit most from anticoagulation.

3 **Investigate PRS for WMH**, including both standard and pathway-based approaches, to elucidate biological mechanisms of cSVD across the lifespan and inform stroke and dementia risk prediction.

This research leverages individual-level clinical and genetic data from large population-based cohorts and biobanks, a memory clinic-based cohort, and the PRESTIGE-AF clinical trial. We validate existing PRS models and develop novel models using integrative and pathway-based frameworks. By applying and refining multiple PRS approaches across key applications, this thesis

seeks to generate critical evidence for the use of PRS in cerebrovascular disease. The findings will support the translation of genomic tools into clinical and public health practice, advancing precision prevention strategies for cerebrovascular disease.

Chapter 3 – Methods

In this thesis, I investigated PRS across various longitudinal cohorts described in the following section, including large-scale biobanks, a memory clinic cohort, and a clinical trial involving survivors of ICH with AF. I will also discuss the longitudinal competing risk models utilized to assess PRS in relation to stroke and dementia. Finally, I will provide a detailed description of constructing PRS, ranging from basic to more complex methods implemented in this work, and how I evaluated PRS using metrics and models relevant to risk prediction and stratification.

In this section I refer to study populations and specific methods as they relate to each of the three core articles in this thesis relevant to each of the three aims.

- **Article 1 (iPRS-DEM):** Integrative PRS for dementia evaluated in elderly community-persons, multi-ancestry biobank, and a memory-clinic cohort.
- **Article 2 (PRESTIGE-AF):** PRS models applied to secondary stroke prevention in ICH survivors with AF.
- **Article 3 (WMH (ps-)PRS):** standard and pathway-specific PRS for WMH, evaluated across the lifespan and in association to stroke and dementia.

3.1 Study Populations

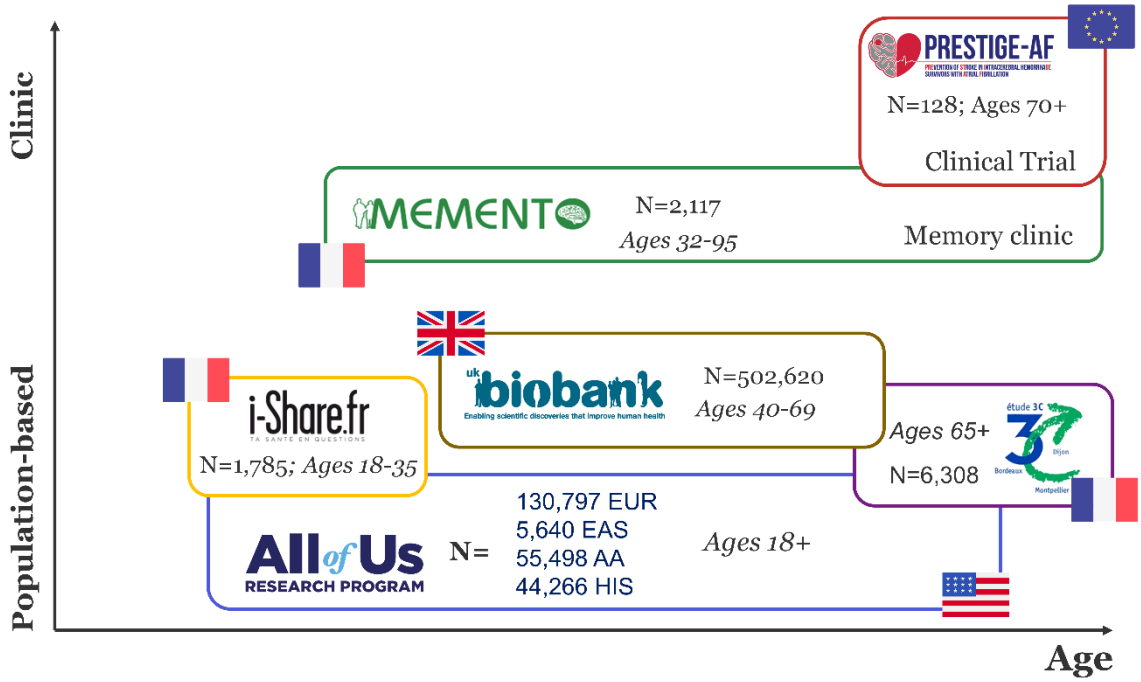


Figure 14. Overview of study populations. In this thesis, I investigate PRS in three French cohorts including the Three City Study (3C) of older community-dwelling persons, the i-Share study of university students, and the Memento memory-clinic study. This work also features the international PRESTIGE-AF phase III clinical trial comprised of intracerebral hemorrhage survivors with atrial fibrillation. I also leverage large-scale biobanks: the UK Biobank, and the All of Us Research Program in the USA.

3.1.1 Three City Study (3C)

3C is a French population-based cohort comprising community-dwelling persons aged ≥65 aiming to evaluate the risk of dementia and cognitive impairment attributable to vascular risk factors in the general population (244). In this thesis, the full 3C cohort with genotyped data is used in Article 1 (iPRS-DEM) and in Article 3 (WMH-psPRS). The MRI subset from 3C-Dijon is also used as a follow-up cohort in Article 3.

Recruitment: Participants were recruited from the electoral rolls of three French cities, Dijon, Montpellier, and Bordeaux between 1999-2001. To be eligible, individuals needed to be non-institutionalized, over 65 years old, and living in one of three cities or their suburbs. Invitations to participate in the study were sent out by post to eligible persons living within selected administrative districts in each city, also inviting their spouse/partner to participate if they met eligibility criteria. The acceptance rate was 37%, resulting in a total of 9,693 persons, after which 399 individuals were either younger than 65 or refused the baseline medical interview. The final, 9,294 participants (2,104 in Bordeaux, 4,931 in Dijon, and 2,259 in Montpellier) underwent baseline assessment. Approval for the study protocol was obtained from the Ethical

Committee of the University Hospital of Kremlin-Bicêtre, and all participants provided informed consent.

Baseline Assessment: In Bordeaux and Dijon participants were first examined at home, undergoing an interview and cognitive testing with trained psychologists or nurses. They were then invited to a study examination center to complete the rest of the protocol – individuals who refused were examined a second time at home. General data collection included demographics, education level, occupational history, monthly income, a short food frequency questionnaire, including detailed information on alcohol and tobacco consumption. Self-reported questionnaires were filed out relating to lifestyle, daily activities, sleep and family history of cardiovascular disease and dementia. Data was also collected related to medical history focusing on cardiovascular diseases, and medication use. Additional assessments were performed to assess cognition, disability level, depression, and screen for Parkinson's disease. The clinical examination collected data on blood pressure and anthropometric measurements, and, for those reporting a history of stroke, examination for sequelae via the Rankin score.

Follow-up: Individuals were followed up every two to three years for up to 19 years. Mortality was ascertained from the civil registry.

Dementia and Stroke Diagnosis: Stroke incidence was recorded at each follow-up examination, participants were asked if they had had a stroke history or stroke symptoms, and if they had been hospitalized. An expert panel of neurologists reviewed potential stroke cases identified at each follow-up visit with participants or informants for deceased participants (245). Diagnosis was made on the basis of all available clinical information, including brain imaging when available, confirming diagnosis of stroke and its subtype (ischemic or haemorrhagic) according diagnostic criterion of the WHO (246). A stroke was defined as a new focal neurologic deficit of sudden or rapid onset, of presumed vascular origin, that lasted 24 hours or more, or leading to death.

Diagnosis of dementia followed a three-step procedure at baseline and at each follow-up visit (244). First, all participants underwent a battery of neuropsychological tests at baseline and each follow-up visit by a trained psychologist, including screening based on the MMSE and Isaacs' Set Test. Second, a neurologist examined all the participants in Montpellier and in Bordeaux at baseline. In Dijon, participants screening positive for dementia based on their neuropsychological performance underwent further clinical examination. At each follow-up, all participants with suspected dementia were examined by a neurologist to establish provisional diagnosis. Third, an independent committee of neurologists confirmed all potential dementia

cases based on the criteria of DSM-IV: having (1) impairment in memory and in at least 1 of aphasia, apraxia, agnosia, or disturbance in executive functioning; (2) impairment and decline in social or occupational function; and (3) cognitive deficits that do not occur exclusively during the course of a delirium episode.

MRI Acquisition: In this thesis, we leverage MRI data only from the 3C-Dijon study (247). In Dijon, participants less than 80 years old were asked if they would undertake an MRI examination (N=2,763). Individuals were excluded if they were carrying internal electrical/magnetic device(s); a history of neurosurgery or aneurysm, claustrophobia, or metal fragments in the eyes, brain or spinal cord. 2,285 participants agreed to participate (82.7%), and only 1,924 examinations were performed due to financial limitations.

MRI scans in 3C-Dijon were acquired from a 1.5-Tesla Magnetom scanner (Siemens, Erlangen, Germany) (247). T1- and T2-weighted images of each subject were aligned to each other using the AIR package. These images were analyzed with the optimized Voxel-Based Morphometry (VBM) protocol, using Statistical Parametric Mapping 99 (SPM99) that were modified in order to consider the structural characteristics of the aged brain.

Genotyping: Ninety percent of the participants accepted to have blood sample for the measurement of biological parameters at baseline. These samples were centralized by the biochemistry Laboratory of the University Hospitals of Dijon. In total, 6,489 participants underwent genome-wide genotyping performed at the Centre National de Génotypage in Evry (France) using the Illumina Human610Iquad BeadChips. Standard quality control procedures were initially conducted on raw genotyped data. This included removing SNPs due to missingness (<98% call rate), low minor allele frequency (MAF) (<1%), non-autosomal location, and deviation from Hardy-Weinberg equilibrium ($p<0.001$). Participants were removed with discordant sex information, excessive missingness (<95% call rate), high heterozygosity (±3SD), or divergent ancestry (±6SD) (248). Additionally, sample pairs were identified with 3rd-degree relatedness with KING software (kinship > 0.0625). The raw genotyped data was then imputed via the Michigan server using the HRC (249).

3.1.2 MEMENTO

The Memento study is a memory clinic-based cohort aimed to investigate the natural history of AD and related disorders in a large group of participants with MCI or subjective cognitive complaints (250). Memento contributed as an external validation cohort in Article 1 and as a

follow-up cohort in Article 3. The memory-clinic represents an important setting to potentially implement PRS for targeted prevention and selection of individuals into clinical trials.

Recruitment: A total of 2,323 participants were recruited from 26 French memory clinics between April 2011 to June 2014. Memory clinics part of the French national network of university-based memory clinics (Centres de Mémoires de Ressources et de Recherche) were asked to participate if they were sufficiently large, had access to a 1.5 or 3-T MRI scanner, and biobank facilities. Participants were screened for either very mild to MCI or isolated cognitive complaints. Very mild to MCI was defined as (1) performing 1-SD worse than an age-sex-education standardized level in one or more cognitive domain on tests performed in the six preceding months, and (2) having a Clinical Dementia Rating ≤ 0.5, and not being demented. For those with subjective cognitive complaints, an individual was eligible if they did not have an objective cognitive deficit based on neuropsychological testing, and were 60 years or older. Subjective cognitive complaints were assessed using visual and analogue scales. The final sample included 370 participants (16%) meeting inclusion criteria for subjective cognitive complaints.

Individuals were excluded if they presented with neurological disease, history of head trauma, previous intracranial surgery, stroke diagnosed in the past 3 months including persistent neurological deficits, or history of schizophrenia. They were also excluded if they had known mutations in familial AD genes, were under guardianship, residence in a skilled nursing facility, or were illiterate. All subjects additionally had to agree to undergo brain MRI, and provide biobank samples. CSF sampling and PET imaging were optional.

Written informed consent was obtained from all participants, and the study protocol received approval from the ethics committee "CPP Sud-Ouest et Outre-Mer III."

Baseline Assessment: Baseline data collection included socio-demographics, personal and family medical history, current medication use, and a clinical examination to collect anthropometric measurements and blood pressure. A battery of neuropsychological tests was administered, and a Clinical Dementia Rating was obtained through an informant. Neuropsychiatric symptoms were also assessed, and data was collected on lifestyle indicators such as smoking and alcohol consumption, physical activity level, leisure activity frequency, and level of daily functioning. Baseline assessment was performed by trained neurologists, with training sessions organized and standardized across study centres.

Follow-up: Follow-up took place every six months for up to 5-years. Socio-demographic characteristics, medical history or incident events, physical examinations, a full neuropsychological battery, and lifestyle questionnaires were administered at each follow-up visit. Biobanking and follow-up brain MRI were conducted at 24 and 48 months, including optional PET scans and lumbar puncture if participants consented.

Dementia diagnosis: In Memento, trained neurologists administered a battery of neuropsychological tests at baseline, and assessed dementia status at each follow-up visit (250). An independent panel of expert neurologists reviewed and validated all possible dementia cases according to DSM-IV criteria. Dementia etiology was ascertained using criteria from the National Institute of Neurological and Communicative Disorders and Stroke and Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) (251).

MRI Acquisition: In Memento, 86% of participants had a 3 or 1.5 Tesla MRI scan at their respective memory clinics (250). Images were harmonized by the Center for Automated Treatment of Images (CATI) (252), and standard acquisitions were applied according to a systematic qualification procedure ensuring parameter uniformity and image quality. MRI sequences were conducted to ensure compatibility with Alzheimer's Disease Neuroimaging Initiative (ADNI) protocol (253).

Genotyping: Biological samples were stored in a centralised biobank (Genomic Analysis Laboratory-Biological Resource Centre, Pasteur Institut Lille)**.** Memento contributed to the EU-JNPD funded European Alzheimer's & Dementia Biobank (EADB) consortium (12). EADB-France coordinated genotyping of all contributing cohorts at the Centre National de Recherche en Génomique Humaine (CNRGH, Evry, France). Samples that passed quality control were genotyped with the Illumina Global Screening Array microarray. Pre-imputation QC included removal of SNPs with MAF<0.01, call-rate<0.98 and HWE<0.001, removal of samples with call-rate <0.05, heterozygosity beyond 3SD, failed sex-check using genotype data of X-chromosome, related sample based on identity by descent (IBD) (>0.1875), or principal component analysis (PCA) outliers beyond 6SD of PC1 and PC2. The raw genotyped data was then imputed via the Michigan server using the HRC reference panel (249).

3.1.3 i-Share

The internet-based Students HeAlth Research Enterprise (i-Share) project, initiated by the University of Bordeaux and University of Versailles Saint-Quentin in France, is a prospective population-based study of university students aged 18-35 years (254). Two ancilliary studies,

MRi-Share and Bio-Share, were also conducted in which a sub-set of participants were invited for brain MRI and genome-wide genotyping (255,256). The MRi-Share study aims to characterize late-maturational changes of the post-adolescent brain and study brain phenotypes and imaging biomarkers of late life disorders, namely cSVD MRI-markers. Bio-Share collected biobank data from blood samples to generate large-scale multi-model molecular biomarkers. In Article 3, the two i-Share ancillary studies are leveraged as a key follow-up cohort to investigate early manifestations of cSVD in young adults.

Recruitment: Participants were eligible if they were officially registered at a University, were at least 18 years of age, and were able to understand French and provide informed consent. Recruitment began in February 2013 and enrollment required registration on the online i-Share portal and the completion of self-administered online questionnaires regarding their physical and mental health, and lifestyle habits. Participants were invited to complete follow-up questionnaires on an annual basis. Between October 2015 and December 2017, at the Bordeaux site, participants were given information about the MRi-Share and Bio-Share sub-studies, and invited to make an appointment with investigators. Participants were screened for exclusion criteria, including (1) over 35 years; (2) pregnant or nursing; (3) claustrophobia; and (4) contraindications for head MRI. Participants then signed informed written consent. A total of 1,870 participants underwent an MRI scan, and 38 scans were removed due to incidental findings, instrument failure, or withdrawal from the study. Nearly all participants in MRi-Share also participated in Bio-Share (N=1,862). The Bio-Share and MRi-Share studies were approved by the regional Ethics Committee (Comité de Protection des Personnes Sud-Ouest et Outre-Mer).

MRI Acquisition: In MRi-Share, MRI scans were acquired with a Siemens 3T Prisma scanner with a 64-channels head coil between November 2015 and November 2017. The MRI protocol was designed to emulate the UK Biobank imaging study, including T1, FLAIR, susceptibility-weighted imaging, diffusion-weighted imaging and resting state functional MRI (257). All 1,832 participants completed the structural scans (T1 and FLAIR), 15 did not complete the diffusion-weighted imaging due to technical problems or anxiety. T1 and FLAIR images were processed with SPM12 (https://www.fil.ion.ucl.ac.uk/spm/) and Freesurfer v6.0 (http://surfer.nmr.mgh.harvard.edu/). For dMRI, pre-processed images were fit to diffusion-tensor imaging modeling, and microstructural model with NODDI (74). A large database of image-derived phenotypes was generated using these images and various analysis pipelines at the brain, regional and voxel level.

Genotyping: Genome-wide genotyping of 1,872 i-Share participants was performed using the Affymetrix Precision Medicine Axiom Array at McGill Genome Center (Canada). After quality control, genotyped data were available for 1,862 participants. Seven participants were removed due to sex discrepancies, two participants who appeared to be duplicates but not twins, and one participant with a kinship coefficient >0.0625 (third degree related) with more than 20 other participants. After applying pre-imputation quality control procedures (SNP call rate <98%, Hardy-Weinberg Equilibrium $p<0.001$), genotypes were imputed on the HRC reference panel.

3.1.4 PRESTIGE-AF

The PRESTIGE-AF trial is a phase-3 randomized control trial (RCT) in six European countries funded by the European Commission as part of the Horizon 2020 research and innovation program (NCT03996772). The trial was designed to investigate the safety and efficacy of DOAC in ICH survivors with AF. The results of the main trial are published in *Lancet* (102). Data collected in the genetic sub-study formed the basis of Article 2.

Recruitment: The PRESTIGE-AF trial recruited 319 individuals, assigning 158 to the DOAC group and 161 to the no anticoagulant (control) group (102). Patients were eligible for recruitment if they were > 18 years older with spontaneous ICH, AF, an indication for anticoagulation, and a modified Rankin score of ≤ 4. Enrollment was 14 days to 12 months after the index ICH. Individuals were excluded if their index ICH was caused by vascular malformation or trauma, or if individuals had or planned to have an implant of a left atrial appendage occlusion device. Written informed consent was obtained from the patient or their legal representative.

At each recruitment site, participants were asked if they wished to take part in the genetic sub-study at the time of main trial recruitment, and again at subsequent follow-up visits until the end of the trial (May 31, 2024). A total of 130 participants (40.8%) were included in the genetic sub-study.

Randomization and Treatment Assignment: Patients were randomly assigned (1:1) to receive either DOAC or no anticoagulation, stratified by location of the index ICH (lobar vs. non-lobar) and sex. Patients, clinicians and investigators were unblinded to treatment assignment. Patients assigned to the DOAC group received apixaban, dabigatran, edoxaban or rivaroxaban. Patients in the no anticoagulation group received antiplatelet medication or no antiplatelet medication as decided by local investigators.

Follow-up: Patients were followed up for at least 6 months and up to 36 months, with in-person follow-up visits at 1, 6, 12, 24, and 36 months. Basic procedures at each follow-up included blood pressure measurements, blood sample collection, and information on adverse events and drug adherence.

Trial Endpoints and Diagnoses: Outcome events were assessed by an event adjudication committee that was masked to treatment allocation, drug use, and participant identity. Co-primary endpoints were incident ischemic stroke and recurrent ICH. An ischemic stroke was defined as an episode of neurological dysfunction caused by focal cerebral, spinal, or retinal infarction lasting ≥ 24 hours. Etiology was determined based on pathological or imaging or clinical evidence of ischemic injury persisting ≥ 24 hours or until death after other etiologies were excluded. Primary ICH was defined as clinical signs of neurological dysfunction attributable to focal collection of blood with the brain parenchyma or ventricular system, with CT/MRI or autopsy evidence, that was not caused by trauma. Secondary endpoints, included, any stroke, major bleeding according to the International Society on Thrombosis and Haemostasis (ISTH) classification (258), and major adverse cardiovascular events (MACE): any stroke, systemic embolism, myocardial infarction, cardiovascular mortality, or major bleeding. Definitions of secondary endpoints can be found in **Appendix 7**.

Genotyping: Samples were collected at their enrolment visit, or at follow-up visits if consent was provided later on, and then centralized at the Vall d'Hebron Research Institute (VHIR) in Barcelona, Spain. DNA was extracted from the buffy coat obtained from blood, and a back-up EDTA tube of whole blood was also drawn. Genome-wide genotyping was performed using the Affymetrix Precision Medicine Axiom Array at the Clinical Research Laboratory and Biobank – Genetic and Molecular Epidemiology Laboratory (CRLB-GMEL, Canada). Pre-imputation was performed, removing variants on both autosomes and chromosome X having a Hardy- Weinberg equilibrium test (p-value<0.001), showing a missingness >0.02 and with a Minor Allele Frequency <0.01. After quality control, genotype data was available for 128 participants (one participant was removed due to sex discrepancy, and one was removed due to excess heterozygosity). Quality control and HRC imputation with the Helmholz Munich server (249) are fully described in **Appendix 7**.

3.1.5 UK Biobank

The UK Biobank (UKBB) is a prospective population-based British cohort of 502,620 participants aged 40 to 69 years at baseline, recruited between 2006 and 2010. The UKBB combines clinical data, including data from hospital registries, biological samples, and imaging

into an extensive biomedical database. This data has been central to the advancement of the field of genetic epidemiology over the past decade in addition to many other scientific discoveries in human health. We were granted access to UKBB data through application number 94113. The UKBB is primarily featured in Article 3, where we leverage brain imaging data, as well as investigate associations with stroke and dementia outcomes. In Article 2 it is used in sensitivity analyses to investigate index event bias.

Recruitment: Between 2006 and 2010, 9.2 million individuals registered with the UK's National Health Service (NHS) were eligible and invited to one of 22 UKB assessment centers across England, Scotland, and Wales.

Baseline Assessment: At baseline, data was collected through self-administered questionnaires, a computer-assisted interview, physical and functional measures, and collection of blood, urine and saliva (259). This covered: socio-demographics; family history and early-life exposures; psychosocial, environmental and lifestyle factors; medical history and current medication use; blood pressure and anthropometrics.

Outcome Ascertainment: Health record linkages were available for all UKBB participants across death, cancer, and hospital inpatient and outpatient registries. These sources rely on codes from international classification of disease (ICD-9 or 10) or Office of Population Censuses and Surveys Classification of Interventions and Procedures (OCPS). Outpatient hospital data is available from 2003 to present, and from 1997 to present for inpatient episodes. Primary care data was also available for 45% of participants up until 2018, which includes medical prescriptions as well.

In this thesis we rely on algorithmically-defined outcomes for stroke and dementia (https://biobank.ndph.ox.ac.uk/showcase/ukb/docs/alg_outcome_main.pdf). The UK Biobank Outcome Adjudication Group, in conjunction with clinical experts, developed and validated algorithms based on the list of clinical codes to provide a standardized approach to defining health outcomes in the UK Biobank. This combines data from the UK Biobank baseline assessment, linked hospital-admission data, and the death registry. The earliest recorded date of a given health outcome is recorded and the source of this data is also provided.

MRI Acquisition: Multi-modal imaging assessment started in 2014, and the UK Biobank has since conducted brain imaging on over 90,000 participants as of March 2025. All surviving UKBB participants were invited via email or postal invitations sent between April 2014 and February 2020, depending on their region. Participants unable to lie still, hold their breath

voluntarily, or hear instructions, as well as those with MRI incompatible metal implants in their body (excluding limbs), were excluded. Details of the UK Biobank MRI protocol and documentation is publicly available (https://www.fmrib.ox.ac.uk/ukbiobank/) and image-derived phenotypes are fully described in a previous publication (257). Briefly, the data were acquired with a standard Siemens Skyra 3T 32-channel head coil. The total scanning time takes 31 minutes to acquire multiple image modalities including: T1-weighted, T2-weighted, susceptibility-weighted, diffusion-weighted, resting-state function al MRI, and task function MRI.

Genotyping: DNA was extracted from stored blood samples from the baseline visit, and genotyping was carried out by Affymetrix Research Services Laboratory. In total, the UK Biobank genotyped 488,377 participants using either the UK BiLEVE Axiom Array by Affymetrix (49,950 participants) or the UK Biobank Axiom Array (438,427 participants). Quality control and imputation with the HRC reference panel are fully described elsewhere (260).

3.1.6 All of Us Research Program

The National Institutes of Health's *All of Us* (AoU) *Research Program,* is a large ongoing prospective cohort study of American adults (aged 18+) aiming to enroll up to 1 million participants and welcoming participants from all backgrounds to reduce disparities in medical research (261). We did not have direct access to AoU data, and rather it was included as part of an external validation (in Article 1) performed by collaborators at Yale School of Medicine. Here I provide a short description of the program that will be one of the most extensive resources in the field in the coming years.

Recruitment: Of note, the program prioritizes recruitment of diverse populations in order to improve diversity in research, particularly in genetics. It has provided funding to a network of 22 community partners focused on engaging diverse communities. Starting in July 2016, participants enroll digitally through the AoU website (https://joinallofus.org). Consenting individuals provide authorization to share electronic health records (EHRs), and complete the initial baseline survey. They are then eligible to undergo evaluation for physical measurements and biospecimen collection at a participating health care provider or through a "direct-volunteer" mechanism (261).

Health Outcome Ascertainment: Health outcomes and conditions, like dementia and stroke, are ascertained from EHR, clinical notes, and self-report. Investigators can identify cases with a combination of Systemized Nomenclature of Medicine (SNOMED) and ICD-9 and 10 codes.

Genotyping: Blood samples (15 mL) were collected from consenting participants for genetic analysis, with saliva samples used as an alternative when blood collection was not possible. As of 2023, WGS had been performed on 245,388 participants who also have linked EHR (262). 46% of these individuals were from under-represented racial and ethnic minorities. Briefly PCR-free barcoded WGS libraries were constructed with the Illumina Kapa HyperPrep kit, and libraries were pooled and sequenced on the Illumina NovaSeq 6000 instrument. Uniformity in genomic data was ensured through a consistent library preparation protocol (PCR-Free Lapa HyperPrep), sequencing on the Illumina NovaSeq 6000 array, and the quality control with the Illumina DRAGEN bioinformatics pipeline. A comprehensive description of these procedures is provided elsewhere (262,263).

3.2 Survival analysis and competing risk models

Much of the work presented in this thesis involves longitudinal data and modeling time-to-event outcomes for incident stroke and dementia. Analyses of this data is within the framework of survival analysis, which requires consideration of censoring (i.e., unknown event times) and the selection of an appropriate time scale.

3.2.1 Choice of time scale

The choice of time scale depends on study design and context. In population-based cohort studies, age is the most natural time scale, as we are interested in the age of disease onset, particularly in elderly populations (264). In clinical studies, especially trials, patients are typically followed from the start of treatment or diagnosis, making time-on-study the most relevant time scale. For the populations analyzed in this thesis, 3C and UKBB were assessed using age as the time scale, while PRESTIGE-AF used time-on-study. In Memento, we occasionally used time since baseline assessment since the progression from cognitive complaints to dementia is of particular interest. Choosing age as the time scale introduces left truncation, meaning subjects are event-free until they enter the study (265). Consequently, individuals contribute to the analysis only at ages for which they were observed, resembling an open cohort design that allows subjects to enter and exit the risk set (266). However, this can cause issues in cohorts like Memento, where small risk sets at younger ages lead to instability in risk estimation (267). The choice of time scale impacts how one controls for age, the strongest

determinant of stroke and dementia risk. Specifically, in regression modeling, using age as a time scale non-parametrically controls for age effects when estimating the baseline hazard function of a Cox proportional hazards model (264). While time-on-study typically controls for age by including it as a continuous covariate at baseline, this may lead to model misspecification if age has a non-linear effect on the outcome that is not accounted for (264). Additional misspecification may arise if there is a significant association between a covariate of interest and age. These considerations were taken into account during my thesis work, and an appropriate time scale was applied whenever possible. Of note, the primary R package, riskRegression (268), used for conducting predictions in the presence of competing risks, particularly in Article 1, did not accommodate left-truncated data, so time-on-study was employed for these analyses.

3.2.2 Censoring

Censoring is a crucial aspect of survival analysis that occurs when the event of interest is not observed for some subjects, resulting in partial or incomplete information about the timing of these events (266,269). Left censoring happens when an event occurs before a subject is included in the study. In the analyses presented in this thesis, left censoring is managed by excluding individuals with pre-existing or prevalent cases of dementia or stroke; thus, we only include individuals who are event-free at the study's outset. Right censoring occurs when the event of interest occurs after the study period ends, or when individuals withdraw or are lost to follow-up. In longitudinal cohorts like 3C and Memento, individuals who have not experienced the event (e.g., dementia) are 'censored' at the date of their last visit. For UKBB, which links hospital data to define disease outcomes, I censor individuals at the last date when hospital data is most complete. In PRESTIGE-AF, most individuals are censored at the trial's conclusion if they did not experience an outcome of interest. Additionally, interval censoring occurs when a subject is observed intermittently (e.g., at follow-up visits), and the event is known to have occurred within that interval. To address interval censoring, a common imputation technique involves using the midpoint between the last observed event-free time and the time the event was recorded. I applied this method for 3C and Memento data, while in UK Biobank, subjects were 'right-point' censored, meaning the date of the observed event was used. These straightforward imputation techniques may introduce some bias, and other more advanced methods have been recommended (266,270). The key with censoring, is to consider whether or not censoring is ‘informative’, in other words, people that were lost to follow-up were lost for reasons independent from the event of interest. Conventional statistical methods rely on this

assumption of non-informative censoring, as if losses to follow-up are at random, and are likely to biased if this assumption is not met (271).

3.2.3 Competing Risks

When analyzing incident stroke and dementia, we often encounter competing risks, the most common being death. A competing event is an event whose occurrence precludes the occurrence of the event of interest. Competing risks is an important violation of the independent censoring assumption, and has significant implications for statistical modeling (271).

The competing risk of death introduces complications with interval censoring, as disease status may be unknown at the time of death (272). In this work, if death occurred before the next scheduled follow-up visit, I often considered the patient 'event-free' at the time of death. However, if they died more than two years after their last event-free visit, they were censored at their last visit. The illness-death model attempts to address the imprecision and bias resulting from these simplified approaches to handling interval censoring and the competing risk of death (272). However, this model was not utilized here due to challenges with implementing prediction in existing R packages.

Most importantly, competing risks bias the standard Cox proportional hazard model (271,273). In standard survival analysis, where death is the outcome, one models the survival function, S(t), representing the probability of survival as a function of time, and reciprocally, the cumulative incidence function (CIF), F(t), representing the probability of dying. We also have the hazard function, h(t) to represent the instantaneous rate of death. In this setting, there is a one-to-one relationship between the hazard function, the survival function, and the CIF.

$$\mathrm{S(t)} = \mathrm{e}^{-\mathrm{H(t)}} \text{ and } \mathrm{F(t)} = 1 - \mathrm{S(t)}$$

- where $H(t)$ is the cumulative hazard or integral of the hazard function, over time T;

$$H(t) = \int_0^{\mathrm{T}} h(\mathrm{t})d\mathrm{t}$$

The most common statistical methods include the Kaplan-Meier method for estimating the survival function and the Cox proportional hazards model for estimating the effects of covariates on the hazard or instantaneous rate of the event of interest (274,275). From Cox proportional hazards models, we can estimate the relative effect, or hazard ratio, of a covariate on the hazard. Due to the established one-to-one relationship, a covariate that increases the hazard (or death rate) will also increase cumulative incidence (or risk of death). However, in the case of competing risks, the one-to-one relationship with the cumulative incidence function

is lost, since the risk of an event also depends on the hazards of all other competing events (271). It is well-established that the Kaplan-Meier estimator results in overestimation of incidence of an event in the presence of competing risks (276). Likewise, the Cox proportional hazard model, does not consider competing risks and may yield biased estimates and poor predictive performance (273,277). Current guidelines strongly recommend competing risk analysis, treating death as the primary competing event, to account for selection bias resulting from differential survival of enrolled participants in dementia research (278,279). Competing risks also arise when analyzing disease sub-types; for example, in the case of first stroke occurrence, one will either experience an ischemic or a hemorrhagic stroke, and the occurrence of the other subtype must be accounted for.

3.2.4 Cause-specific hazard vs. Fine-Gray Regression models

In this work, I employed the cause-specific hazard (CSH) model and the Fine-Gray Regression (FGR) model, the primary methods used to model competing risks (271,280). Both approaches account for competing risks, but define their risk sets differently and use different hazard functions for each type of event. This has important implications for interpreting regression coefficients in each model, and modeling the CIF.

The CSH model can be implemented by treating individuals who experience a competing event as censored at the time of that event, effectively removing them from the risk set. Essentially, one censors at the competing event, creating a 'cause-specific' hazard model for each event type. The CSH model therefore estimates the instantaneous rate (hazard) of events among subjects who are currently event-free. This allows for interpretable cause-specific hazard ratios for covariates estimated in the model among those who survived all events up to time *t*, making this approach suitable for etiological questions (281). However, the CIF, and therefore risk, can only be accurately modeled by the CSH model under the untestable assumption that competing risks are independent.

FGR models are alternatively widely used to model associations with cumulative incidence (i.e. risk). The FGR model, also called the sub-distribution hazard model, defines the instantaneous risk of failure for a specific event among subjects who have not yet experienced that event, including subjects in the risk set who have experienced other competing events. In other words, we consider participants who have experienced the competing event (death) as part of the risk set, even if they contribute ‘immortal’ time (i.e. experienced the competing event). This is

necessary to maintain an explicit one-to-one link to the cumulative incidence function, implying that if a covariate increases the sub-distribution hazard, it will also increase the event's incidence. This approach is recommended for prediction as it overcomes limitations of Cox proportional hazard models in estimating the effect of covariates on the CIF (271). However, FGR coefficients, which correspond to sub-distribution hazard ratios, require careful interpretation. While the direction of association can be described, the magnitude of this effect on the CIF cannot be directly quantified. One can only make statements on relative magnitude of effects between different covariates on the incidence of the same outcome within the same study (282). The drawback, from an etiological standpoint, is that regression coefficients are not directly interpretable, and in fact are frequently misinterpreted (282). Even so, simulations have shown that sub-distribution hazard ratios may approximate the effect on the risk of the event when the probability of the event is relatively infrequent (<20%) (282).

Overall, it is recommended to apply a CSH model when the focus is etiological due to the interpretability of covariate coefficients, while a FGR model is preferred when the objective is to estimate risk or predict prognosis (279,282). By utilizing both modeling approaches, one can more comprehensively explore the impact of competing events on associations between covariates and the rate of events (with CSH) or the probability of the event over time (with FGR). In this work, I often consider the CSH model as the primary analysis and the FGR model as a sensitivity analysis when focusing on associations between PRS and incident stroke or dementia—essentially addressing the etiological question of whether genetic risk affects stroke or dementia incidence. However, when prediction is the goal, as in Article 1, I employ an FGR model that better facilitates the determination of the effect of PRS (e.g., iPRS-DEM) on dementia risk over time.

3.3 Polygenic Risk Scores

The methodology for calculating and evaluating PRS involves a systematic pipeline that transforms genetic association data into a standardized risk estimate of genetic risk burden for a specific trait or disease. Prior to constructing PRS the raw genotyped data and GWAS summary statistics need to follow careful quality control procedures, described in **Appendix 3**.

3.3.1 Principals of constructing PRS

Most complex traits and common diseases are polygenic and under the influence of hundreds, thousands, or millions of SNPs that, individually, only exert small effects on disease risk. However, when combined together in a PRS, they can capture genetic predisposition to a trait

or disease. At a fundamental level, a PRS is the weighted sum of the number of risk alleles across a set of SNPs, using effect sizes typically estimated in GWAS for a given phenotype. A PRS for a given individual, *i*, can be represented as:

$$PRS_i = \sum_{j=1}^{m} \beta_j X_{ij}$$

Where:

- m is the total number of SNPs in the PRS model
- β_j is the effect size for the *j-th* SNP (extracted from GWAS)
- X_{i_j} is the number of effect alleles for *i*-th individual's *j*-th SNP.

The summation of many small effects of common SNPs make it so that in a population of sufficient size PRS can be expected to follow a normal distribution with higher percentiles indicating higher genetic predisposition (**Figure 15**). A PRS, by condensing genetic information into a single value, can then be used as a continuous variable representing an individual-level estimate of genetic risk for a trait.

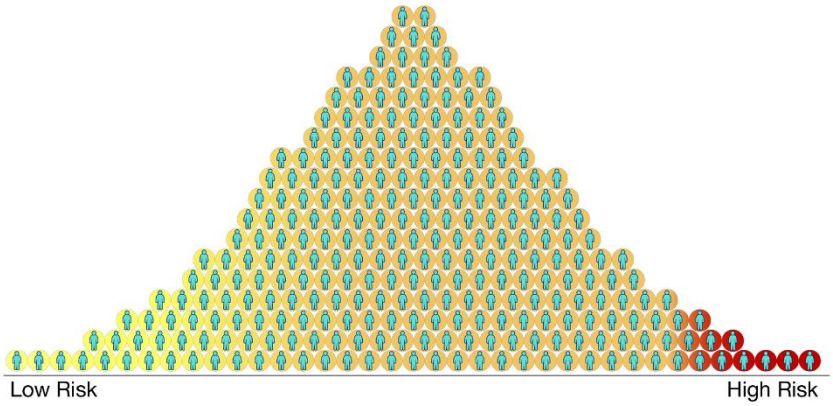


Figure 15. PRS follow a normal distribution in a given population of appropriate sample size. Often those at the extreme of the distribution are designated as high risk. *Adapted from the National Human Genome Research Institute.*

PRS methods must also account for two core statistical principles: independence and imprecise estimates. Independence is necessary to avoid multicollinearity. Most correlation observed among SNPs is due to LD. If LD is not accounted for, we would be essentially double-counting non-independent variants resulting in a biased PRS. Furthermore, patterns of LD can vary markedly across populations which influences effect size estimation and can limit the

portability of PRS across ancestry groups (237). Quality control procedures and advanced statistical approaches are designed to deliberately address LD in PRS construction.

Imprecise effect estimates also introduce noise into the estimation and can greatly reduce prediction accuracy. Imprecision arises from the complexity of genetic architecture, and insufficient sample size of the underlying GWAS. SNPs that occur at lower frequency (<1%) in the population are much more susceptible to imprecision. In addition to excluding rare SNPs, modern PRS methods employ different shrinkage techniques, 'shrinking' imprecise effect estimates towards zero, aiming to optimize the signal-to-noise ratio to yield the best predictive power.

3.3.2 Standard PRS Methods

Numerous methods have been developed to construct PRS that aim to tackle LD and effect shrinkage in different ways. Here I will briefly discuss validated PRS methods used in this work though many other PRS methods also exist (**Appendix 4**).

3.3.2.1 Genome-Wide Significant Variants (Genomic Risk Scores)

The earliest and simplest approach to construct PRS using only variants that are significantly associated to a trait at the GWS level ($p<5x10^{-8}$) (283). Traditionally, these were termed genetic risk scores (GRS) but are now often interchangeably called PRS. Here, I will use GRS to refer to basic scores that are limited to only these variants. These scores may be advantageous due to their intuitive appeal, that the score only included markers that are robust, and therefore would be more easily conveyed to clinicians and policy makers (284). Yet, for many traits, these GRS have limited predictive ability and it has been consistently shown that prediction is optimized by including many additional variants beyond GWS (208,284). Still, for their simplicity and interpretability, a GRS approach represents a useful starting point to benchmark the added value of more complex PRS including up to millions of SNPs.

3.3.2.2 Clumping + Thresholding

The most widely used 'traditional' PRS method is called Clumping and thresholding (C+T) (285). In this approach, SNPs are first "clumped", ideally against an external reference panel, to include only independent SNPs so that LD is accounted for. Clumping is an iterative process in which the most significant SNP in a specified genetic distance, *w*, is selected, and any variants correlated with this index variant beyond a chosen r^2 threshold are removed. Next a p-value threshold, p_t, is applied to shrink effect sizes to zero for the clumped SNPs that have a p-value over the specified threshold.

3.3.2.3 Bayesian Methods

C+T has noted limitations, namely that information is lost in the clumping process where potentially causal SNPs are discarded, and effect sizes set to zero based on a strict cut-off resulting in limited prediction accuracy (286). Ideally, PRS would be able to jointly model all SNPs across the genome to make full use of available information (287). Two Bayesian approaches to PRS, *LDPred2* and *PRS-CS* software attempt to limit loss of information by directly modeling LD from an external reference panel (e.g., the UK Biobank LD reference panel or 1000 Genomes) (287,288). These Bayesian methods, though substantially more computationally expensive, have shown to improve prediction over C+T (287–289).

Briefly, LDpred2 derives PRS based only on an LD matrix (286,288). LDpred2 requires the estimation of two parameters to compute the LD-matrix; the fraction of causal SNPs (ρ), and SNP heritability (h^2_{SNP}) estimated from the GWAS summary statistics with LD score regression (170). PRS-cs is a similar approach distinguished by utilizing a continuous shrinkage prior on SNP effect sizes (287). PRS-cs requires tuning of the global shrinkage parameter (ϕ) that represents prior knowledge about the sparseness of the underlying genetic architecture. More recently, PRS-CS authors have also developed an approach to improve multi-ancestry PRS prediction (PRS-CSx) (290).

3.3.2.4 PGS Catalog

While in this work I use a combination of these Bayesian methods and C+T to construct PRS, researchers are also highly encouraged to upload PRS models on the PGS catalog (https://www.pgscatalog.org/) (291). The PGS catalog attempts to address reproducibility in PRS research, standardize PRS reporting, and promote PRS translation. It is an immense resource which currently has over 5,000 PRS for 660 traits across 750 publication, enabling researchers to apply previously validated PRS models to their independent datasets. This can be of particular interest when researchers lack a dataset to optimize PRS for a given trait. Ideally, for each trait and given population we can determine the best and most generalizable, and assessing the transportability of existing PRS models should be considered in most studies. As such, I use existing models from the PGS catalog in Article 2 to directly apply existing PRS models to clinical trial data.

3.3.3 Advanced PRS methods

Many other PRS methods have been developed leveraging different statistical approaches, use additional data sources such as functional annotations or multiple ancestry reference panels, and have different use cases (**Appendix 4**). In this section, I introduce a strategy to combine

PRS for multiple traits in an integrative PRS framework (featured in Article 1) and also a method to deconstruct PRS into biologically relevant pathways in Article 3.

3.3.3.1 Integrative PRS

Because numerous PRS methods exist, each capable of generating multiple models, prediction accuracy may be improved by systematically optimizing and integrating PRS across these methods. Typically, PRS are integrated through penalized regression framework that selects the optimal linear combination of PRS. Indeed, combining PRS models derived for a single-trait, has shown to improve prediction (285,292,293). Applied to a particular disease, a "meta-genomic" PRS was developed for CAD combining three previous PRS for CAD into one meta-genomic score with over 1.7 million SNPs (293). This integrative score showed substantially greater association with CAD risk as well as improved prediction compared to the individual PRS scores and conventional risk factors (293). Methods and software have been developed to streamline and optimize integration across multiple PRS models. For example, a 'stacked' C+T approach identifies the optimal linear combination of multiple C+T models and weights their contributions to the final, 'stacked' model accordingly, and showed substantial prediction improvement over any single C+T PRS (285). This has been extended in PRSmix, which, using elastic net penalized regression, aggregates genetic information for all PRS published on the PGS catalog which includes various PRS methods (e.g., PRS-CS, LDpred-2, etc.) that were previously optimized in diverse cohorts (292). Across numerous traits, compared to the best PGS catalog PRS, PRSmix improved mean prediction accuracy by 1.2-fold in both Europeans and South Asians (292).

There is also potential of a multi-trait PRS approach to improve cross-trait prediction (285,292,294–296) taking advantage of shared genetic architecture and high genetic correlation across many traits, and that many genetic variants exhibit pleiotropy and may influence disease through intermediate endophenotypes (211,297). This approach may be of particular value for heterogenous, complex conditions with many contributing risk factors and likely underlying causes. This approach has been applied to ischemic stroke, first in 2019, combining existing ischemic stroke PRS with PRS for multiple risk factors and comorbidities like blood pressure and lipid traits and diabetes (211). A later study, went further including PRS derived from cross-ancestry stroke GWAS and showed added benefit beyond clinical risk factors in biobanks of Europeans and East Asians, and across five clinical trials (10). A similar approach has also been done for ICH (212), showing improvement in prediction as well.

This approach requires at least one independent 'training' dataset to derive weights for each single-trait PRS included in the iPRS. These weights are typically extracted from penalized regression, such as elastic net regression (298) that accounts for correlation between PRS by shrinking coefficients and performing sparse variable selection. Notably, by deriving weights for single-trait PRS, one can identify which traits contribute most to prediction of the overall trait and also construct a per-SNP-weighted iPRS that can be transported and validated in other datasets. The per-SNP weighted score follows the general formula:

$$iPRS_i \propto \sum_{j=1}^{m} X_{i_j}\left(\frac{\gamma_1}{\sigma_1}\alpha_{j_1} + \cdots + \frac{\gamma_n}{\sigma_n}\alpha_{j_n}\right)$$

where *m* is the total number of SNPs in the iPRS, $\sigma_1, \ldots, \sigma_n$ are the empirical standard deviations of each of the *n-th* sPRS in the training set, $\alpha_{j_1}, \ldots, \alpha_{j_n}$ are the effect size for the j-th SNP in each PRS model respectively, and X_{i_j} is the number of risk alleles for *i*-th individual's *j*-th SNP. A SNP's effect size α_{jn} was equal to zero in a given PRS if the SNP was not present. A full description of this methodology applied to stroke is available in the original publications, and I detail an adaptation of this methodology in Article 1.

This approach to integrating information across PRS also forms the basis of published, scalable methods, PRSmix+ and multi-PGS (292,294). In PRSmix+, all the well-powered PRS available in the PGS catalog (over 2,600) are inputted into an elastic net regression – including those for traits other than the target trait—and mixing weights are used to aggregate effects in the same framework described above (292). Importantly, they showed the value of adding cross-trait PRS, leveraging pleiotropic effects, and demonstrated improvement in prediction performance over PRSmix which optimizes using only trait-specific PRS. Multi-PGS, rather than using the PGS catalog, applied the auto-tuning implementation of LDpred-2 across 937 publicly available GWAS to build a "PRS library" (294). In this approach, in addition lasso penalized regression, XGBoost (299), a non-linear method, is used to derive training weights for each PRS and covariates. They uniquely show that their method can improve prediction accuracy for a range of diseases including those for which there is no available GWAS summary statistics.

The core trade-off between implementing highly automated iPRS pipelines and developing more disease-specific integrative models is optimizing for prediction accuracy versus retaining control over the inputs and maintaining model interpretability. For instance, automated approaches like Multi-PGS or PRSmix+, which combine hundreds or thousands of PRS, inherently dilute the model's interpretability by combining many potentially correlated traits

regardless of whether they truly share genetic architecture contributing to the target disease. This contrasts with targeted, disease-specific models, such as the approach developed for ischemic stroke, where researchers deliberately selected PRSs for vascular risk factors (e.g., blood pressure and lipids) and known comorbidities. It is possible, that even though these methods perform variable selection and regularization procedures, the over-inclusion of unnecessary PRS influences the final selection of PRS in the model. Rather than comparing against existing single-trait PRS, iPRS methods should be compared where selection on input traits is made based on prior knowledge compared to automated procedures, but this has not been done to my knowledge.

3.3.3.2 Pathway-specific PRS

Standard PRS methods discussed thus far aim to provide a single, global genetic risk estimate, and improve disease prediction. Meanwhile pathway-specific PRS (ps-PRS) partition genetic liability into biologically meaningful pathways allowing us to better clarify the etiological heterogeneity of complex diseases given the functional sub-structure of the genome (**Figure 16**). Essentially, pathway-based PRS may be more useful for investigating disease etiology, and clarifying disease subtypes. Here I highlight a ps-PRS methodology implemented in PRSet that is central to Article 3.

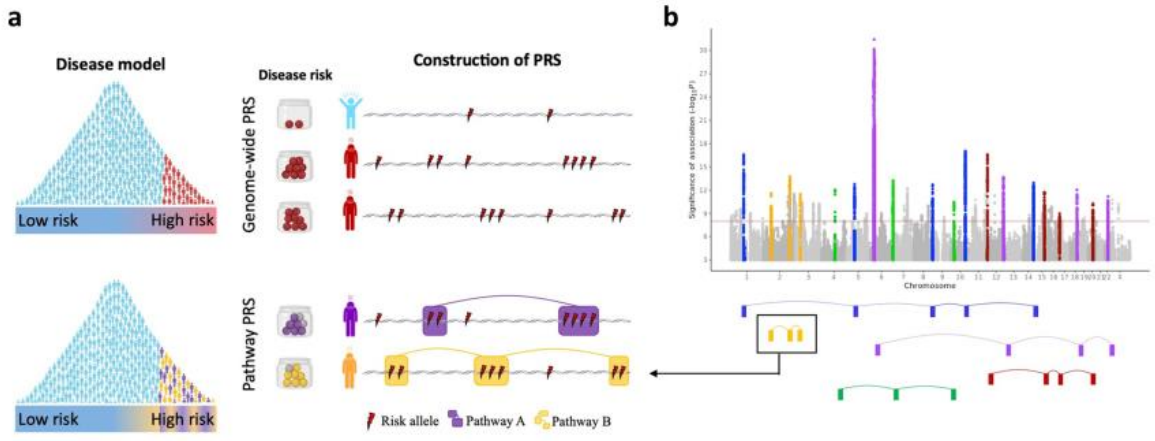


Figure 16. The 'pathway polygenic' risk score approach is described here. Colors represent genes within the same genomic pathway. The main takeaway here is that while a genome-wide PRS captures global genetic risk, it is likely that different pathways contribute to overall risk, and individuals will have different susceptibility to these pathways. A pathway-PRS is able to characterize an individual's susceptibility to underlying pathways that drive their genetic risk to a disease. *From Choi et al. 2023*.

First, ps-PRS can be defined with a priori knowledge of pathological processes, as was initially done for AD (300), defining pathways for Aβ clearance or cholesterol metabolism. More recently, ps-PRS have been defined through the use of canonical databases representing biochemical knowledge (301–304), or using data-driven approaches with pathway enrichment

tools like MAGMA (305). This hypothesis-free approach has uncovered enriched pathways involved in other conditions such as diabetes (306–308), hypertension (309), Parkinson's disease (310), and multiple sclerosis (311). Ps-PRS have also uncovered pathways predisposing to AD-related changes in brain structure (224), and CSF biomarker and early imaging endophenotypes (312).

Recently, PRSet, implemented within PRSice-2 (v.2.3.5) (313) was developed for efficiently computing and analysing pathway-based PRSs, in which PRS are calculated across genomic pathways for each individual (314). To use PRSet, one needs the GWAS for the phenotype of interest, genotyped data from a target dataset, and additionally a GMT file to define gene sets and a GTF file to define gene boundaries to determine if a SNP falls within a specific gene. Naturally, it leverages the classical C+T approach in PRSice-2 to calculate *k* PRS corresponding to *k* genomic pathways for an individual, *i*:

$$PRS_{ik} = \sum_{j=1}^{m_k} \beta_j X_{ij}$$

where m_k is the number of clumped SNPs in pathway *k*, β_j is the SNP effect size estimated from a GWAS on the studied phenotype, and X_{ij} is the genotype of individual *i* for SNP *j*. The clumping procedure is slightly different than for the genome-wide C+T method. PRSet performs clumping on each pathway independently in order to retain pathway signal and account for correlation between SNPs in nearby genes of the same pathway. This also ensures that SNPs present in multiple pathways are indeed counted in each.

The PRSet authors demonstrated the utility of PRSet for pathway enrichment, showing comparative performance to standard pathway enrichment software used in GWAS like MAGMA (305) and LDSC (315). Ps-PRS has the distinct advantage of directly quantifying an individual's genetic susceptibility within specific biological pathways, and these ps-PRS can uniquely be used in association testing with other outcomes and in independent cohorts.

For pathway enrichment, PRSet relies on empirical, "competitive" p-values derived from a permutation-based procedure. PRSet is applied on clumped SNPs (r^2<0.1) without any p-value thresholding. This is to ensure that all genes in the pathway are represented regardless of their underlying association to the trait. The permutation-based procedure also accounts for the number of SNPs included in a given pathway, so as not to privilege pathway enrichment towards pathways with a larger number of genes.

In brief, the method is summarized in the PRSet paper:

1. A “background” pathway containing all genic SNPs is constructed, and clumping is performed within this pathway. For pathways with m SNPs, N null pathways are generated by randomly selecting m “independent” SNPs from the “background” pathway.
2. The competitive P-value can then be calculated as

$$competitive\ P-value = \frac{\sum_{n=1}^{N} I(P_n < P_o) + 1}{N + 1}$$

Where: I(.) is an indicator function, taking a value of 1 if the association p-value of the observed gene set (P0) is larger than the one obtained from the nth null set (Pn), and 0 otherwise. A pseudo-count of 1 is added to the numerator and denominator to avoid competitive P-values of 0 and conservatively counting the observed gene set as 1 potential null set. One consideration of this permutation procedure is that the smallest achievable competitive P-value is 1/(N+1), which can lead to difficulties in ranking highly significant gene sets.

In Article 3, I primarily use PRSet to identifying underlying biological pathways for WMH, and their connection to early life markers of white matter microstructure, and to late life cSVD complications like stroke and dementia. Ps-PRS may also be specifically useful for disease sub-typing and risk stratification among a range of other applications, and represents a novel and promising methodology complementing genome-wide PRS.

3.3.4 Optimizing PRS

Optimizing PRS is similar to any prediction paradigm. Ideally, a PRS should be optimized in a training dataset and then validated in a testing dataset. If an independent cohort of sufficient size is available, then it can be split into a training and testing set (usually 80/20%). Alternatively, two independent cohorts can be used for training and testing separately. The training set is used to select the "best" PRS model from a set of candidates PRS models constructed based on different PRS methods and hyperparameters. The best PRS model is then evaluated in the testing dataset.

In C+T, for example, this includes selecting the optimal clumping (r^2) and p-value thresholds – often a range r^2 values ranging from 0.01 to 0.5 are tested, and for p_t thresholds are tested from genome-wide significance ($p<5x10^{-8}$) to including all clumped SNPs ($p<1$). Bayesian methods have their own set of hyperparameters as mentioned above: LDpred-2 parameterizes the fraction of causal SNPs (ρ), and h^2_{SNP} and PRS-CS has a global shrinkage parameter based on sparseness of genetic architecture, Φ. Of note, both LDpred-2 and PRS-CS, among other methods, have an 'auto-learning' feature for parameter estimation, avoiding the need for an independent training cohort. Still, auto-learning methods based solely on GWAS summary statistics are sensitive to subtle mismatches between the summary statistics and reference panel resulting in misspecification and convergence issues in LD modeling (316). These issues are exacerbated when using a lower-powered GWAS with substantial heterogeneity in SNP effect sizes, in which case an independent dataset should be used to select optimal parameters. Essentially, if an independent tuning cohort of sufficient sample size and power for the trait or disease of interest is readily available, it should be leveraged to optimize PRS.

Another important note is that the choice of training dataset will naturally optimize PRS specific to its context. Especially for integrative frameworks like PRSmix+, a tuning cohort will be used to determine the weights of each individual PRS. Indeed, simulations with PRSmix+ showed that lower heritability traits benefited most from PRSmix+ but also required a greater training sample size to accurately estimate mixing weights (292). Generally, larger training sample size leads to improved prediction compared to smaller samples (289). However, it can be difficult to find independent cohorts, especially since large-scale GWAS meta-analysis currently include many large population-based cohorts with genetic data, including the UKBB. In such cases, the training cohort can be excluded from the base GWAS, but this does reduce statistical power. Researchers should carefully consider these trade-offs and if they have access to a suitable training cohort and an additional 'testing' set as well.

3.3.4.1 Statistical models and evaluation metrics

Once the design has been set, selecting which candidate PRS models will be generated and designating training and testing datasets, the next step is to select the statistical models and evaluation metrics that will be used to optimize and evaluate PRS. The selection of the statistical model depends on the target trait and the type of data available (longitudinal or cross-sectional). Briefly, PRSs are typically evaluated using regression models that test the association between the score and the phenotype/outcome in the target population. The most common models used to evaluate PRS are:

- **Linear Regression:** Used for **quantitative traits** (e.g., height, cholesterol levels) to estimate the per-unit change in the trait due to a per-unit change in the PRS.
- **Logistic Regression:** Used for **binary traits** (e.g., case/control status for a disease) to estimate the odds ratio (OR).
- **Survival Analysis:** The most popular model is the Cox proportional hazard model, used for time-to-event outcomes (e.g., dementia onset) to estimate Hazard Ratios (HRs)**.** Often, in survival analysis, modeling also needs to account for competing risks (e.g., death), in which case cause-specific Cox models or Fine-Gray regression models are often used (271,280).

These models should additionally be adjusted for important covariates, which primarily include demographic variables, such as age and sex, and genetic principal components (PCs). PCs control for population stratification or hidden population structure potentially related to geographical or participation biases within the target cohort. The first 10 PCs are routinely considered, although the number of PCs can be selected based on visuals such as the Scree plot. Often, study-specific variables like genotyping chip used or the recruitment center are often included as covariates or stratification factors to account for technical differences.

While one can look at measures of association (coefficients) in regression models to evaluate PRS, such as hazard or odds ratios, prediction models should be evaluated for their ability to estimate the trait or disease risk for new individuals. Model evaluation involves assessing both discrimination and calibration. Discrimination is the ability of the model to assign higher risk estimates to patients who will experience the event compared to those who will not, while calibration compares the model's predicted probabilities to the observed probabilities. Prediction metrics used in this thesis include the root-mean-squared error for quantitative traits, and area under the curve (AUC) and Brier Score for binary traits. For time-to-event outcomes time-dependent AUC, Brier Score, the C-index, the scaled Brier Score (i.e., the index of prediction

accuracy) and the integrated Brier Score were used. A detailed description of these metrics is included in **Appendix 5.**

Lastly, PRS models are often optimized within an internal validation framework. Internal validation involves estimation of a model's prediction performance on new, unobserved data using the same dataset from which the model was developed. This is opposed to apparent performance, which is the estimation of the model performance on the same data used to develop the model, and this is almost always severely 'optimistic' or over-estimated (317). Often internal validation is done by splitting a dataset into training and testing subsets, however, this "split-sample" approach is inefficient and results in loss of power and could lead to unstable prediction estimates (318). More efficient procedures include bootstrapping and k-fold cross validation.

- **Bootstrapping:** Bootstrapping involves drawing random bootstrap samples with replacement from the original development data set. Prediction performance (e.g., AUC or Brier Score) is assessed and averaged over each bootstrap sample. Often many bootstrap resamples are needed to stabilize prediction metrics (M>200).
- **K-fold cross-validation:** K-fold cross-validation involves partitioning the training set into k parts. The performance is estimated by predicting the observations in each part using a model trained on the remaining k-1 parts of the data. Often k=10 or k=5 is used.

These internal validation approaches are used particularly in Article 1 in attempt to optimize PRS and attain unbiased prediction metrics. PRS models are also externally validated in other populations, where we are more concerned about the prediction model's generalizability (319).

3.3.5 Investigating PRS Associations

PRS generally aims to improve overall risk prediction and prognosis, in which case the above prediction metrics are useful. Beyond prediction, association studies can give more context for PRS use in risk stratification. We should also consider different ways of assessing and parameterizing PRS, including in the context of environmental and clinical factors, and whether genetic effects may be modulated by other relevant variables.

When various PRS are calculated their absolute values can be of different scales or ranges depending on how many SNPs are included in the score and the direction of effect alleles. Thus, PRS is most frequently standardized to a normal distribution (mean = 0, and SD = 1) in order to compare PRS and to allow for combination of PRS in integrative frameworks. In this case, the hazard ratio or odds ratio is interpreted as per 1 SD increase in PRS. Continuous PRS values can also be visualized by comparing cases versus controls (**Figure 17**).

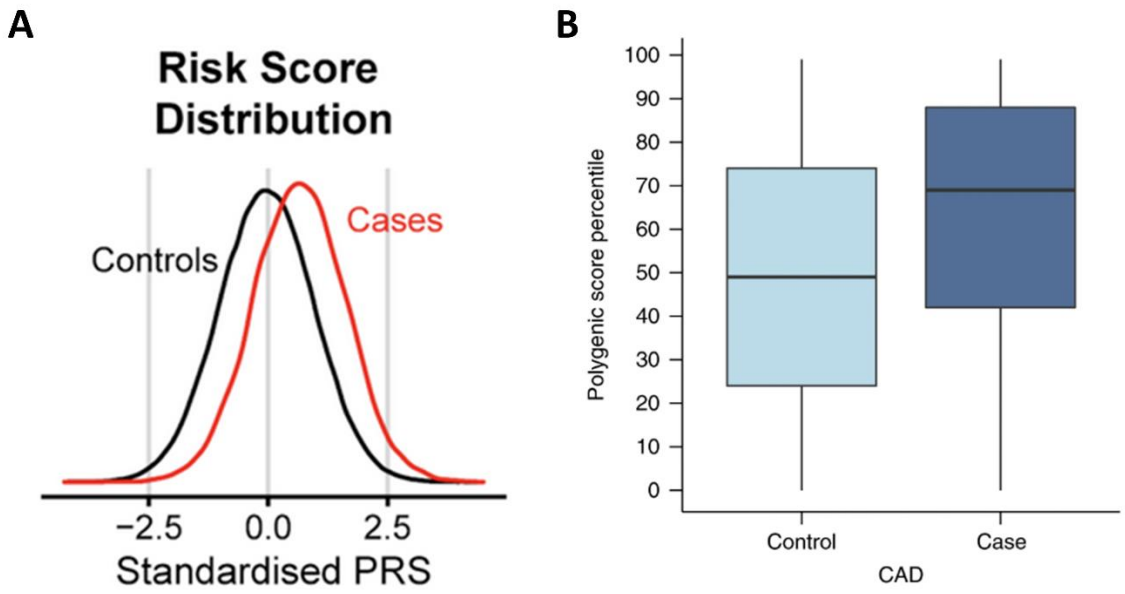


Figure 17. Visualizing continuous PRS in cases versus controls (A) comparing standardized distributions of PRS cases vs. controls, adapted from Wand et al. (2024). (B) Box plot visualization of continuous PRS for CAD, adapted from Khera et al. (2018) (208).

Often PRS is assessed as a categorical variable, which can help to make results more easily interpretable for risk stratification-- allowing one to designate "high" versus "low" PRS for example. Indeed, it has often been shown that those at the extreme of a PRS distribution have much higher risk compared to those with average PRS values (208,211). There are different ways to categorize PRS:

1) **Arbitrary Quantiles:** Arbitrarily dividing the PRS into quantiles like tertiles, quartiles, quintiles, or deciles (if sample size permits), and comparing to a reference group, often those with the lowest PRS or average PRS values (e.g., lower or middle tertiles).
2) **Extremes of Distribution:** Since polygenic risk may only be concentrated in the extremes of a PRS for some diseases, often it is useful to compare the top 10%, 5% , and 1% compared to the lower or middle risk stratum. While this can be very useful for clinical interpretation and communication, statistically, we restrict to only a subset of the sample and association measures can be quite imprecise without large enough samples (**Figure 18A**).
3) **Sliding Window Approach:** This approach involves comparing 'high' versus 'low' PRS using different percentile cut-offs (e.g., top 5% versus bottom 95%) (**Figure 18B**). This allows us to keep the whole sample, and to see if the risk increases linearly across the PRS distribution. This approach can be used to define cut-offs according to a pre-defined relative risk threshold (e.g., OR > 2).

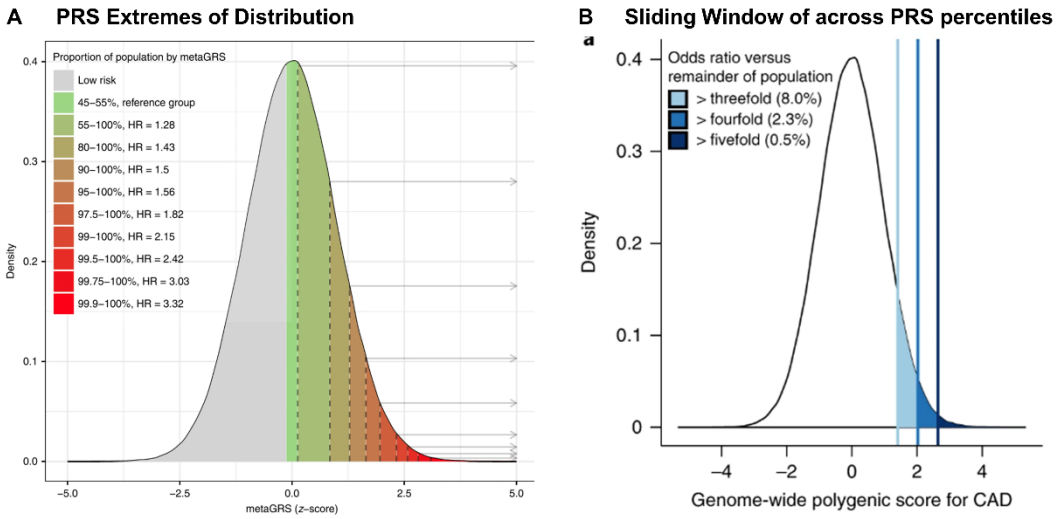


Figure 18. Two common ways to categorize PRS. **(A)** *Adapted from Abraham et al. (2019).* This shows the comparison for an ischemic stroke metaGRS between an 'average' risk reference group (45-55% PRS decile) compared to individuals into the extreme of PRS distribution (up to top 0.1%). **(B)** *Adapted from Khera et al. 2018*. A CAD PRS is compared at increasing percentile cut-offs versus the rest of the population. The authors show percentile cut-offs where the odds ratio represents a 3-fold (top 8%), 4-fold (top 2.3%), and 5-fold risk (top 0.5%) increase versus the remainder of the population.

By categorizing PRS in these ways we can also visualize relative risk across the PRS distribution, and we can further calculate and visualize absolute risk over time (i.e., cumulative risk) across different PRS quantiles and sub-groups. Indeed, these investigations are vital for determining genetic risk thresholds for different diseases for their eventual implementation in clinical risk prediction models (320).

3.4.5.1 Investigating PRS in context of other genetic, clinical, and environmental risk factors

PRS should be investigated in the context of relevant clinical, environmental, and possibly other genetic factors relevant to the disease of interest. Primarily, one could test if the PRS is associated with the disease *independent of* established risk factors. This is done by simply adjusting for a set of additional risk factors in an appropriate regression model. Further, it is common practice to compare the added value of PRS, above and beyond risk factors, in a prediction setting. Models including established risk factors independently and combined can be compared to models including PRS only and models including both genetic and non-genetic risk factors.

It is also of interest if genetic risk captured by PRS is modulated by other pertinent variables, like sex, hypertension status, or notable genetic risk variants not included in the PRS (e.g. *APOE* in AD). Interaction, or effect modification, can be directly tested in regression models, and strata-specific relative and absolute risks can also be compared. A common strategy is to investigate different PRS quantiles within clinically-defined strata. For example, previous studies

investigating PRS for stroke have stratified by those with different levels of risk factors (hypertension, BMI, smoking) (211), or those on antithrombotic medication (212).

In the context of dementia, in Article 1, I performed numerous analyses that investigated iPRS-DEM alongside *APOE* genotype, and established risk factors (e.g., age, sex, diabetes, low education, obesity, history of cardiovascular disease). These analyses, described in detail in Chapter 4, spanned prediction performance, interaction testing, and adjustment for the aforementioned variables. In Article 2, analyses reflect particular investigation into if DOACs are more effective at reducing risk in individuals of high or low genetic risk (as defined by various PRS), and improvement in prediction of trial endpoints beyond clinical risk scores. In Article 3, we investigate if hypertension status modulates the association between WMH ps-PRS and WMH volume to identify underlying mechanisms of hypertension-driven cSVD.

In addition, I employed a factorial design in both Articles 1 and 2 to further investigate clinically-relevant risk stratification. This approach involved creating groups defined by genetic risk (PRS quantiles) and an additional variable of interest where we were particularly interested in effect modification (**Figure 19**). Relative and absolute risk can be compared between these groups. For example, in Article 1, we particularly create groups based on *APOE-ε4* carrier status and high vs. low PRS, allowing us to investigate the ability of PRS to identify individuals at high genetic risk beyond *APOE*. We can also assess the magnitude of genetic risk relative to a reference group consisting of those without *APOE-ε4* and below the high-PRS cut-off (e.g., bottom 90%) representing the majority of the general population. In Article 2, we use a similar approach to define groups based on their treatment assignment (treatment or control group) and PRS tertile (lower vs. upper). Comparing these strata allows us to determine if DOACs seem to be more effective in those in higher or lower genetic risk determined by a given PRS. These investigations are quite valuable for determining the potential clinical utility of PRS applied to different prevention contexts.

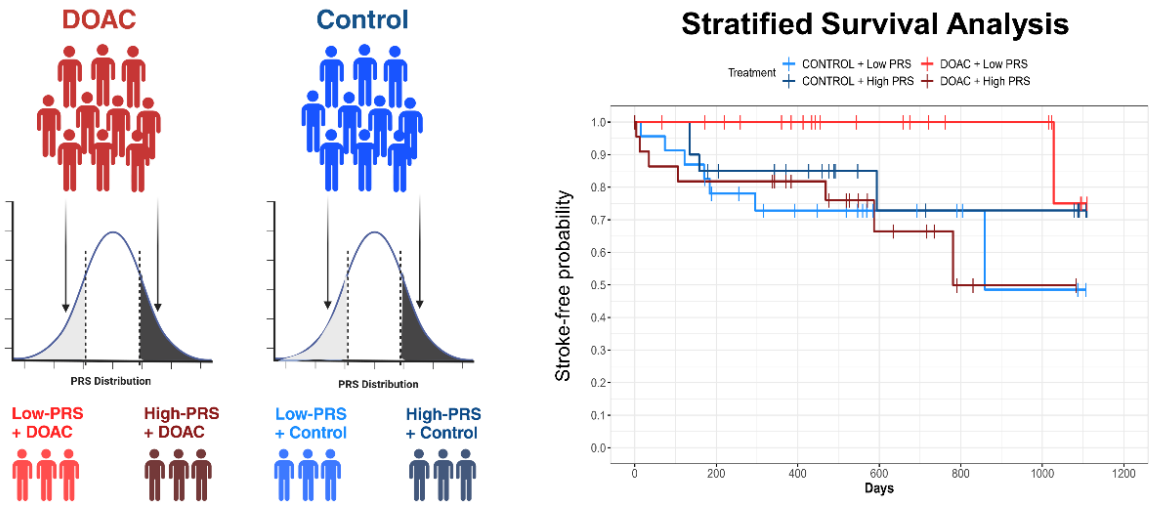


Figure 19. Factorial design to stratify into clinically-relevant groups. This example is from the study design of Article 2, where we stratify based on treatment arm and high vs. low PRS. We can test if DOAC reduces stroke risk more effectively in individuals in high vs. low PRS (comparing dark or light-colored groups). Similarly, we can see if PRS may modulate DOAC effectiveness (comparing light vs. dark red groups).

Chapter 4 – Integrative PRS for dementia

The first project of my thesis was a continuation of work I had started during my Master's internship.

4.1 Polygenic score integrating neurodegenerative and vascular risk informs dementia risk stratification

We start with PRS applied in a primary prevention setting with the overall aim to improve dementia risk prediction by developing an integrative PRS that better captures both the vascular and neurodegenerative burden of dementia. As explained previously, dementia is a clinical syndrome characterized by cognitive deficits that impact daily function, and its pathology is very heterogenous. Dementia patients often have multiple underlying pathologies, and it is becoming increasingly accepted most dementia cases involve a "mixed" pathology, predominantly driven by vascular and neurodegenerative processes (4,5). From a primary prevention standpoint, it may be more advantageous to determine an individual's all-cause dementia risk, especially since mid-life risk factors make up a large percentage of the modifiable risk of dementia. Multimodal interventions, including management of vascular risk factors during midlife, have shown promise in preventing dementia, particularly among high-risk groups (138,140,321).

Compared to a standard PRS, derived from individual traits to primarily predict risk of that given trait (e.g, AD-PRS for AD), an integrative approach could capture additional genetic risk specifically acting through intermediate endophenotypes and risk factors, and thus improve prediction and refine risk stratification. This approach is inspired by previous work on an integrative PRS for ischemic stroke (211), that was also applied in the GIGASTROKE project (10), led by my lab in Bordeaux. For AD and dementia, a number of PRS have been developed, which provide evidence for additional genetic risk beyond *APOE* genotype (216,217,221). An integrative AD-PRS or "metaGRS" study was also recently published, but did not show improvements over a standard AD-PRS (222). AD-PRS in these studies often were derived and evaluated in case-control studies specifically designed to study AD pathology, and may neglect other forms of dementia. The AD-metaGRS also did not focus on vascular traits and rather included PRS for traits more globally linked to AD (e.g., social isolation) nor did they evaluate beyond standard logistic regression and prediction metrics. On the whole, these AD-PRS studies do not directly incorporate vascular risk, largely do not consider an all-cause dementia phenotype,

and have not evaluated PRS in longitudinal cohorts while importantly account for competing risks like death.

We aim to address these gaps by integrating PRS for both vascular and neurodegenerative traits, and training PRS models in a longitudinal French cohort of older community-persons while accounting for competing risk of death. We then sought to evaluate the iPRS-DEM in multiple relevant settings: in the general elderly population (3C-Dijon, > 65 years), in the memory-clinic (Memento), and across multiple ancestries in a large-scale biobank (AoU). Further, in both a population-based and memory-clinic setting, we performed a set of analyses in order to investigate potential of iPRS-DEM in prediction beyond *APOE* and clinical risk factors, and also, importantly, we used a factorial design to stratify individuals into genetic risk groups based on iPRS-DEM and *APOE* which could better inform risk stratification.

We used the 3C Bordeaux and Montpellier subset of the 3C study to train iPRS-DEM. To select traits for inclusion in the iPRS-DEM, we identified both vascular and neurodegenerative traits relating to dementia and with available large-scale GWAS. These included 27 PRS for AD (excluding the *APOE* region), vascular risk factors, stroke sub-types, cSVD MRI-markers, and biomarkers of neurodegeneration. The methodology to generate iPRS-DEM involved:

1) Selection of the best single-trait PRS model from 148 candidate models generated across state-of-the art PRS methods. Selection was based on the ability of PRS to predict risk of all-cause dementia over time.
2) Fine-Gray elastic-net penalized regression to derive weights given to each PRS in the model.
3) Integration of PRS by reweighting each SNP in iPRS-DEM

This methodology included important improvements over previous approaches, notably the use of FGR models which are recommended for risk prediction in longitudinal settings with competing risks.

We first highlight that iPRS-DEM was associated with all-cause dementia beyond *APOE* genotype in 3C-Dijon, Memento, and in Europeans and East Asians in AoU. Importantly, in 3C-Dijon, a standard AD-PRS was not associated with all-cause dementia. In terms of overall prediction, iPRS-DEM did not show much improvement beyond *APOE*. However, iPRS-DEM + *APOE* prediction performance was equivalent to individual clinical risk factors (obesity, low education, history of cardiovascular disease) in a population-based setting, and substantially improved prediction in the memory-clinic. Interestingly, we found iPRS-DEM was perhaps most useful in a risk stratification

setting alongside *APOE-ε4* status. In 3C-Dijon, compared to a reference group of non *APOE-ε4* carriers with low iPRS-DEM representing the majority of the general population, those with high iPRS-DEM conveyed nearly a 2-fold increased risk of dementia over time irrespective of *APOE-ε4* status. This stratification was even stronger in the memory-clinic where *APOE-ε4* was a particularly strong predictor. Compared to the reference group, those with high iPRS-DEM and *APOE-ε4* reaching a 5-fold increased risk, while those with *APOE-ε4* but low iPRS-DEM had a roughly 2.5-fold increased risk.

Overall, we show that iPRS-DEM has potential to improve dementia risk prediction and stratification in addition to, and alongside *APOE* genotype. We also importantly show generalizability to independent population-based cohorts, and the memory-clinic. This has important implications for primary prevention to identify those at higher risk of all-cause dementia in the general population and recommend them for primary prevention programs that control lifestyle and vascular risk factors, or select individuals for clinical trial enrichment. We also suggest translatability of iPRS-DEM to East Asians, but this requires further validation, and there is a need to refine iPRS-DEM in other ancestries.

Article 1 (first author)

Tim D'Aoust MSc[1], Santiago Clocchiatti-Tuozzo MD MHS[2,3], Cyprien A. Rivier MD MS[2,3], Aniket Mishra PhD[1], Tsuyoshi Hachiya PhD[4], Benjamin Grenier-Boley[5], Aïcha Soumaré PhD[1], Marie-Gabrielle Duperron MD,PhD[1,6], Quentin Le Grand PhD[1], Vincent Bouteloup MSc[1,7], Cécile Proust-Lima PhD[1,7], Cécilia Samieri PhD[1], Jeanne Neuffer PhD[1], Muralidharan Sargurupremraj PhD[8], Geneviève Chêne PhD[1,9], Catherine Helmer PhD[1], Mura Thibault PhD[10], Philippe Amouyel MD,PhD[5,11], Jean-Charles Lambert PhD[5], Yoichiro Kamatani MD PhD[12], Hélène Jacqmin-Gadda PhD[1], David-Alexandre Tregouët PhD[1], Michael Inouye PhD[13,14], Carole Dufouil PhD[1,7], Guido J. Falcone MD ScD MPH[2,3], Stéphanie Debette MD PhD[1,6],

Contribution of the PhD candidate:

- All analyses apart from the external validation in *All of Us* in which case a detailed analysis plan was provided to the analyst.
- Interpretation of the results
- Writing of the full manuscript

I would particularly like to thank Guido Falcone's team at Yale School of Medicine for their contribution of AoU data to this work during the revision stage. This article is published in *Alzheimer's and Dementia:*

D'Aoust T, Clocchiatti-Tuozzo S, Rivier CA, et al. Polygenic score integrating neurodegenerative and vascular risk informs dementia risk stratification. Alzheimer's Dement. 2025; 21:e70014. https://doi.org/10.1002/alz.70014

The published article is included in the following pages. The supplementary material is available in **Appendix 6**.

Check for updates

Received: 2 August 2024 | Revised: 24 January 2025 | Accepted: 25 January 2025

DOI: 10.1002/alz.70014

Alzheimer's & Dementia®
THE JOURNAL OF THE ALZHEIMER'S ASSOCIATION

RESEARCH ARTICLE

Polygenic score integrating neurodegenerative and vascular risk informs dementia risk stratification

Tim D'Aoust[1] | Santiago Clocchiatti-Tuozzo[2,3] | Cyprien A. Rivier[2,3] | Aniket Mishra[1] | Tsuyoshi Hachiya[4] | Benjamin Grenier-Boley[5] | Aïcha Soumaré[1] | Marie-Gabrielle Duperron[1,6] | Quentin Le Grand[1] | Vincent Bouteloup[1,7] | Cécile Proust-Lima[1,7] | Cécilia Samieri[1] | Jeanne Neuffer[1] | Muralidharan Sargurupremraj[8] | Geneviève Chêne[1,9] | Catherine Helmer[1] | Mura Thibault[10] | Philippe Amouyel[5,11] | Jean-Charles Lambert[5] | Yoichiro Kamatani[12] | Hélène Jacqmin-Gadda[1] | David-Alexandre Tregouët[1] | Michael Inouye[13,14] | Carole Dufouil[1,7] | Guido J. Falcone[2,3] | Stéphanie Debette[1,6,15]

[1]Bordeaux Population Health Center, INSERM, UMR U1219, University of Bordeaux, Bordeaux, France

[2]Department of Neurology, Yale School of Medicine, New Haven, Connecticut, USA

[3]Center for Brain and Mind Health, Yale School of Medicine, New Haven, Connecticut, USA

[4]Iwate Tohoku Medical Megabank Organization, Iwate Medical University, Yahaba, Iwate, Japan

[5]U1167-RID-AGE facteurs de risque et déterminants moléculaires des maladies liées au vieillissement, INSERM, Institut Pasteur de Lille, CHU Lille, University of Lille, Lille, France

[6]Department of Neurology, Institute for Neurodegenerative Diseases, CHU de Bordeaux, Bordeaux, France

[7]CIC 1401 EC, Pôle Santé ´Publique, INSERM, Bordeaux, France

[8]Biggs Institute for Alzheimer's and Neurodegenerative Diseases, University of Texas-San Antonio, San Antonio, Texas, USA

[9]Department of Public Health, CHU de Bordeaux, Bordeaux, France

[10]Institute for Neurosciences of Montpellier INM, INSERM, UMR U1298, University of Montpellier, Montpellier, France

[11]Department of Epidemiology and Public Health, CHRU de Lille, Lille, France

[12]Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Chiba, Japan

[13]Department of Public Health, Cambridge University, Cambridge, UK

[14]Baker Heart and Diabetes Institute, Melbourne, Victoria, Australia

[15]Institut du Cerveau (ICM), Paris Brain Institute, INSERM U1127, UMR CNRS 7225 Paris, Sorbonne Université, Assistance Publique des Hôpitaux de Paris, Paris, France

Correspondence
Stéphanie Debette, Bordeaux University, 146 rue Léo Saignat, Bordeaux Cedex 33076, France; Institut du Cerveau (ICM), Sorbonne Université, 47/83 boulevard de l'Hôpital, Paris 75013, France.
Email: stephanie.debette@u-bordeaux.fr; stephanie.debette@icm-institute.fr

Abstract

INTRODUCTION: An integrative polygenic risk score (iPRS) capturing the neurodegenerative and vascular contribution to dementia could identify high-risk individuals and improve risk prediction.

METHODS: We developed an iPRS for dementia (iPRS-DEM) in Europeans (aged 65+), comprising genetic risk for Alzheimer's disease (AD) and 23 vascular or

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2025 The Author(s). *Alzheimer's & Dementia* published by Wiley Periodicals LLC on behalf of Alzheimer's Association.

Funding information
French National Research Agency, Grant/Award Numbers: ANR-18-RHUS-0002, 11-LABX-0009; France 2030 IHU3, Grant/Award Number: ANR-23-IAHU-0001; Fondation Claude Pompidou; French Investments for the Future program, Grant/Award Number: 17-EURE-0019; Fondation Plan Alzheimer, Grant/Award Number: FCS2009-2012; French Ministry of Research, Grant/Award Number: DGRI 2020-2024; Fondation pour la Recherche Médicale; Caisse Nationale Maladie des Travailleurs Salariés; Direction Générale de la Santé; Institut de la Longévité; Conseils Régionaux of Aquitaine and Bourgogne; Fondation de France; Ministry of Research-INSERM Programme "Cohortes et collections de données biologiques"; National Institutes of Health, Grant/Award Numbers: R01HG010001, 32AG019134, P30AG021342; Prix NRJ-neurosciences Académie des Sciences; Plan Maladies Neurodégénératives; EU Joint Programme - Neurodegenerative Disease Research (BRIDGET project); Mutuelle Générale de l'Education Nationale (MGEN); Prix Burrus, Fondation pour la Recherche Médicale; European Union's Horizon 2020 research and innovation programme under grant agreement No 754517; Prix Fondation Claude Pompidou

neurodegenerative traits (excluding apolipoprotein E [*APOE*]). iPRS-DEM was evaluated across cohorts comprising older community-dwelling people (N = 3702), a multi-ancestry biobank (N = 130,797 Europeans; 105,404 non-Europeans), and dementia-free memory clinic participants (N = 2032).

RESULTS: iPRS-DEM was associated with dementia risk independently of *APOE* in the elderly (subdistribution hazard ratio [sHR]$_{per1SD}$ = 1.15, 95% confidence interval [CI]: 1.03 to 1.28), which generalized to Europeans (EUR-sHR$_{per1SD}$ = 1.28, 95% CI: 1.09 to 1.51]), East-Asians (EAS-sHR$_{per1SD}$ = 5.29, 95% CI: 1.43 to 34.36), and memory-clinic participants (sHR$_{per1SD}$ = 1.25, 95% CI: 1.11 to 1.42). Prediction was comparable to clinical risk factors in older community-dwelling people, with improved performance among memory-clinic patients. Risk stratification was enhanced by defining four genetic risk groups with iPRS-DEM and *APOE* ε4, reaching five-fold increased risk in *APOE* ε4+/iPRS-DEM+ memory-clinic participants.

DISCUSSION: Alongside *APOE* ε4, iPRS-DEM may refine risk stratification for the enrichment of dementia clinical trials and prevention programs.

KEYWORDS
apolipoprotein E genotype, community-dwelling elderly, competing risk analysis, dementia prevention, incident dementia, longitudinal study, memory clinic, multi-ancestry biobank, polygenic risk score, transportability of PRS, vascular cognitive impairment

Highlights

- iPRS-DEM reflects neurodegenerative and vascular contribution to dementia.
- We show iPRS-DEM captures additional dementia genetic risk beyond *APOE* and AD-PRS.
- iPRS-DEM, in combination with *APOE* ε4, shows promise for dementia risk stratification.
- Our results generalize across both population-based and memory-clinic settings.
- We show transportability of iPRS-DEM to East Asian ancestry.

1 | BACKGROUND

Dementia, arising from a complex interplay of genetic and environmental factors, is a major contributor to disability and dependency among older persons, affecting over 55 million people worldwide.[1] Most dementia cases in the population exhibit "mixed" pathology, with co-occurrence of neurodegenerative (predominantly due to Alzheimer's disease [AD]) and vascular brain injury, including those due to clinical stroke and, more frequently, covert cerebral small vessel disease (cSVD).[2–5] This vascular contribution has been sparsely accounted for in dementia trials, and the failure to develop disease-modifying therapies has been partly attributed to over-reliance on imprecise AD biomarkers used to enroll at-risk patients.[6] Yet, approximately 40% of dementia burden is attributable to modifiable risk factors, many of vascular origin (including hypertension, physical inactivity, obesity, smoking, and diabetes).[7] Multimodal interventions, including management of vascular risk factors in midlife, have shown promise in preventing dementia, particularly among high-risk groups.[8–10]

Polygenic risk scores (PRS), derived from large genome-wide association studies (GWAS), capture genetic risk for complex conditions by combining the effects of many independent risk variants on a given phenotype. They have a demonstrated ability to identify individuals at a high genetic risk of multifactorial diseases, with effect sizes approaching those of monogenic disease mutations in the top percentiles of PRS.[11] As PRS can be measured at low cost and early in life, they could be important tools to tailor preventive interventions for high-risk individuals[12,13] and for predictive and prognostic enrichment of clinical trials.[14] PRS derived from AD GWAS (AD-PRS) that combine the effects of common genetic risk variants other than apolipoprotein E epsilon 4 (*APOE* ε4) were shown to predict AD beyond *APOE* ε4 and are associated with neuroimaging markers of neurodegeneration and amyloid and tau burden.[15,16] However, while some studies have described AD-PRS to also predict all-cause and vascular dementia,[17–21] recruitment in AD GWAS is likely biased toward patients with less extensive contribution of vascular brain disease than in the general population.[22]

15525279, 2025, 3, Downloaded from https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.70014 by Cochrane France, Wiley Online Library on [21/10/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

We sought to more comprehensively capture polygenic risk of all-cause dementia by applying a novel integrative PRS (iPRS) approach, combining AD-PRS with PRS for stroke, MRI markers of cSVD, vascular risk factors, and neurodegenerative processes. The iPRS model is based on the assumption that the majority of genetic variants exert their effects on a given disease by affecting intermediate traits and was successfully applied to coronary artery disease (CAD) and stroke.[23–26] A similar approach was recently applied to AD,[27] but this included a heterogeneous set of traits (eg, PRS for psychiatric conditions and sensory deficits) and did not comprehensively capture genetic susceptibility to vascular brain disease. Moreover, published AD-PRS were typically trained and evaluated in case–control datasets and have not explicitly investigated association with the probability of dementia over time (ie, cumulative incidence) while accounting for the competing risk of death.

Here, we trained an iPRS for all-cause dementia (iPRS-DEM), excluding the *APOE* locus, in a population-based, longitudinal cohort of older community-dwelling persons. Using competing risk models, we characterized the relationship of iPRS-DEM with cumulative incidence of all-cause dementia and evaluated the predictive performance of iPRS-DEM against single-trait AD-PRS, in comparison and combination with *APOE* and clinical risk factors. We explored the ability of iPRS-DEM to stratify dementia risk alongside *APOE* ε4 carrier status and validated it in memory clinic patients with cognitive complaints or mild cognitive impairment (MCI) to inform utility of iPRS-DEM to different stages of dementia prevention. Finally, we investigated the transportability of iPRS-DEM in a large, prospective cohort comprising participants of diverse socioeconomic background and ancestries (Europeans, East Asian, African-American, and Hispanic).

RESEARCH IN CONTEXT

1. **Systematic review**: Using PUBMED and similar sources, we identified studies investigating dementia PRS. There were limited studies demonstrating PRS association with all-cause dementia in longitudinal cohorts. Additionally, PRS studies focused primarily on AD and have yet to incorporate genetics reflecting the vascular contribution to dementia risk.
2. **Interpretation**: In longitudinal population-based and memory-clinic cohorts, we found an integrative PRS comprising multiple vascular and neurodegenerative traits, beyond AD-PRS alone, to be associated with risk of all-cause dementia independent of *APOE* and clinical risk factors and to refine stratification of dementia risk. An integrative PRS for all-cause dementia may capture additional risk above and beyond AD polygenic risk, *APOE*, and clinical risk factors and has potential to identify high-risk individuals in multiple dementia prevention settings.
3. **Future directions**: Future studies integrating PRS with clinical prediction models and further development of dementia PRS in non-Europeans are warranted.

2 | METHODS

2.1 | Study populations

The 3C and Memento studies were used to train and validate the iPRS-DEM. 3C is a population-based cohort comprising community-dwelling people aged ≥ 65 years from the French cities of Dijon, Montpellier, and Bordeaux, recruited in 1999 to 2001 and followed every 2 to 3 years over 12 to 17 years.[28] We split 3C by center into a training (Bordeaux and Montpellier [3C-BM]) and validation set (3C-Dijon). Memento participants were recruited from 26 French memory clinics based on subjective cognitive complaints or MCI in 2011 to 2014, with a follow-up every 6 to 12 months over 5 years.[29] The National Institutes of Health's All of Us (AoU) is a large ongoing prospective cohort study of American adults (aged 18+) aiming to enroll up to one million participants and welcoming participants from all backgrounds to reduce disparities in medical research.[30] Among those who consent, AoU collects survey questionnaires capturing detailed past and present medical history data, electronic health records, physical measurements (ie, blood pressure, height, weight), and genomic information (ie, whole genome sequencing).[31] Data were available from all participants who enrolled from the beginning of the program in May 2018 to June 2022 (release version 7). Participants without viable genome-wide genotype data, with prevalent dementia at baseline, or no follow-up were excluded. Genome-wide genotyping and imputation are described in the eMethods 1.

2.2 | Dementia diagnosis

In 3C, diagnosis of dementia followed a three-step procedure at baseline and at each follow-up visit.[28] First, all participants underwent a battery of neuropsychological tests at baseline and each follow-up visit by a trained psychologist. Second, a neurologist examined all the participants in Montpellier and in Bordeaux at baseline. In Dijon, participants screening positive for dementia based on their neuropsychological performance underwent further clinical examination. At each follow-up, all participants with suspected dementia were examined by a neurologist to establish provisional diagnosis. In Memento, trained neurologists administered a battery of neuropsychological tests at baseline and assessed dementia status at each follow-up visit.[29] In both cohorts, an independent panel of expert neurologists reviewed and validated all possible dementia cases according to the *Diagnostic and Statistical Manual of Mental Disorders-Fourth Edition* criteria. In AoU, dementia was ascertained from individual-level electronic health records, clinical notes, and self-report data using a combination of Systematized Nomenclature of Medicine (SNOMED)[32] and International Statistical Classification of Diseases and Related Health Problems, Ninth and Tenth Revisions (ICD-9 and 10)[33] codes shown in eTable 1.

15525279, 2025, 3, Downloaded from https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.70014 by Cochrane France, Wiley Online Library on [21/10/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.3 | Generation of iPRS

An overview of the derivation and validation of iPRS-DEM is given in Figure 1. We identified large-scale GWAS summary statistics for 27 dementia-related traits aiming to comprehensively capture both neurodegenerative and vascular contribution to dementia, including AD, stroke (subtypes), vascular risk factors, MRI markers of cSVD, hippocampal volume, and cerebrospinal fluid levels of phosphorylated tau and amyloid beta (references for included GWAS in eTable 2). We performed standard quality control recommended for PRS,[34] including addressing sample overlap of 3C or Memento in GWAS meta-analyses (eMethods 1). As recommended, the *APOE* region (chromosome 19: 44.4 to 46.5 Mb) was excluded from the AD-PRS.[35]

In 3C-BM, we generated 148 candidate PRS models per trait using three validated PRS methods: the "clumping and thresholding" approach[36] and two Bayesian approaches, LDPRED2[37] and PRS-CS[38] (eMethods 2). Each candidate PRS model was evaluated in a Fine-Gray regression (FGR) model for cumulative incidence of all-cause dementia with time since baseline (years) as a time scale and considering death as a competing event.[39] Models were adjusted for age at baseline, age squared (to account for non-linearity), study center, sex, the first 10 genetic principal components (PCs) of population stratification,[40] and dosage of *APOE ε4* and *APOE ε2* alleles. Dementia onset was imputed as the midpoint between the time of diagnosis and last dementia-free visit. Non-demented individuals, including those who died more than two and a half years since their last visit, were censored at their last dementia-free visit. We performed a systematic procedure to select and integrate the best PRS for each trait and construct iPRS-DEM (Figure 1, eMethods 3). Briefly, we evaluated each candidate PRS model based on the lowest integrated Brier score[41] across 5, 7.5, and 10 years and then inputed the best PRS for each trait in Fine-Gray elastic-net regression[42] to address overfitting and correlation between PRS using the glmnet R package[43,44] (eMethods 3). Training weights were extracted from elastic-net for each sPRS and used to re-weight SNP effect sizes in the iPRS-DEM (eMethods 3).

2.4 | Validation of iPRS-DEM

We estimated the effect of iPRS-DEM and AD-PRS (included in iPRS-DEM) per 1 SD increment on the cumulative incidence of dementia (referred to interchangeably as dementia risk) in 3C-Dijon using FGR and the same time scale, censoring, and adjustment variables as for 3C-BM. We also partitioned each PRS into quintiles, estimating subdistribution hazard ratios [sHRs] relative to the middle 20%, and ran sensitivity analyses with cause-specific hazard (Cox) models for dementia and death.[45] As complementary analyses, we assessed dichotomized versions of each PRS at progressive percentile cutoffs against the rest of the sample. We verified if iPRS-DEM was associated with dementia risk independently of clinical risk factors (hypertension, smoking, dyslipidemia, diabetes, obesity, history of cardiovascular disease, and low education; definitions in eMethods 4) and calculated the percentage change in the FGR coefficient after adjustment (removing those with missing data; n = 26). Of note, FGR coefficients correspond to sHR, which only approximate the magnitude of effect on the cumulative incidence of dementia[46] (eMethods 5). Lastly, we performed subgroup analyses of the association between iPRS-DEM and dementia risk, stratifying by sex and looking at differential associations depending on age at onset (<80, ≥80 years, eMethods 6).

We also assessed risk stratification ability of iPRS-DEM alongside *APOE ε4* after removing participants with *APOE ε2/ε4* genotypes (n = 70). We first tested the interaction of iPRS-DEM with *APOE ε4* carrier status and obtained *APOE ε4* strata-specific sHRs. Then, through a factorial design, we categorized individuals into four genetic risk groups based on their *APOE ε4* status (*APOE ε4+/APOE ε4–*) and the presence of high iPRS-DEM (iPRS-DEM+/iPRS-DEM–) (eMethods 7 for threshold selection). In non-parametric analysis, we compared the cumulative incidence of dementia at different ages in each genetic risk stratum, accounting for death as a competing event and left truncation with the etm R package (etmCIF function)[47,48] (eMethods 8). We also estimated the sHR in each genetic risk stratum relative to non-*APOE ε4* carriers with low iPRS-DEM (*APOE ε4–/*iPRS-DEM–) as the reference group.

Finally, in FGR models, we compared prediction performance of iPRS-DEM against AD-PRS, *APOE ε4* and *ε2* dosage, clinical risk factors (showing significant association with increased dementia risk), and a reference model containing age at baseline, sex, and 10 PCs (eMethods 9). We calculated the time-dependent area under the curve (AUC)[49] and Index of Prediction Accuracy (IPA),[50] a rescaling of the Brier score against a null model with no covariates, at 10-year follow-up over 2000 bootstrap replications in 3C-Dijon using the riskRegression R package.[51] Twenty-six individuals with missing clinical risk factor data were also removed from these analyses.

2.5 | External validation in Memento

The iPRS-DEM was generated in Memento and validated by applying the same methods as described previously. FGR models were similar to 3C-Dijon validation, except for including only a linear term for age at baseline. We investigated approximately similar risk factors in Memento, including family history of dementia (eMethods 10). To estimate cumulative incidence in genetic risk strata, we used time since baseline assessment, adjusted for baseline covariates (adjustedCurves R package[52]) since progression from cognitive complaints or MCI to dementia is of interest in this clinical population (eMethods 11). Again, we performed complete case analysis for analyses including clinical risk factors and removed those with *APOE ε2/ε4* genotype (n = 39) from *APOE ε4* stratified analyses (eFigure 1). All statistical analyses were carried out in R version 4.1.0.[53]

2.6 | Transportability of iPRS-DEM across ancestries in AoU

We investigated the transportability of iPRS-DEM across ancestries in the US population-based All of Us research program. Here,

15525279, 2025, 3, Downloaded from https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.70014 by Cochrane France, Wiley Online Library on [21/10/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

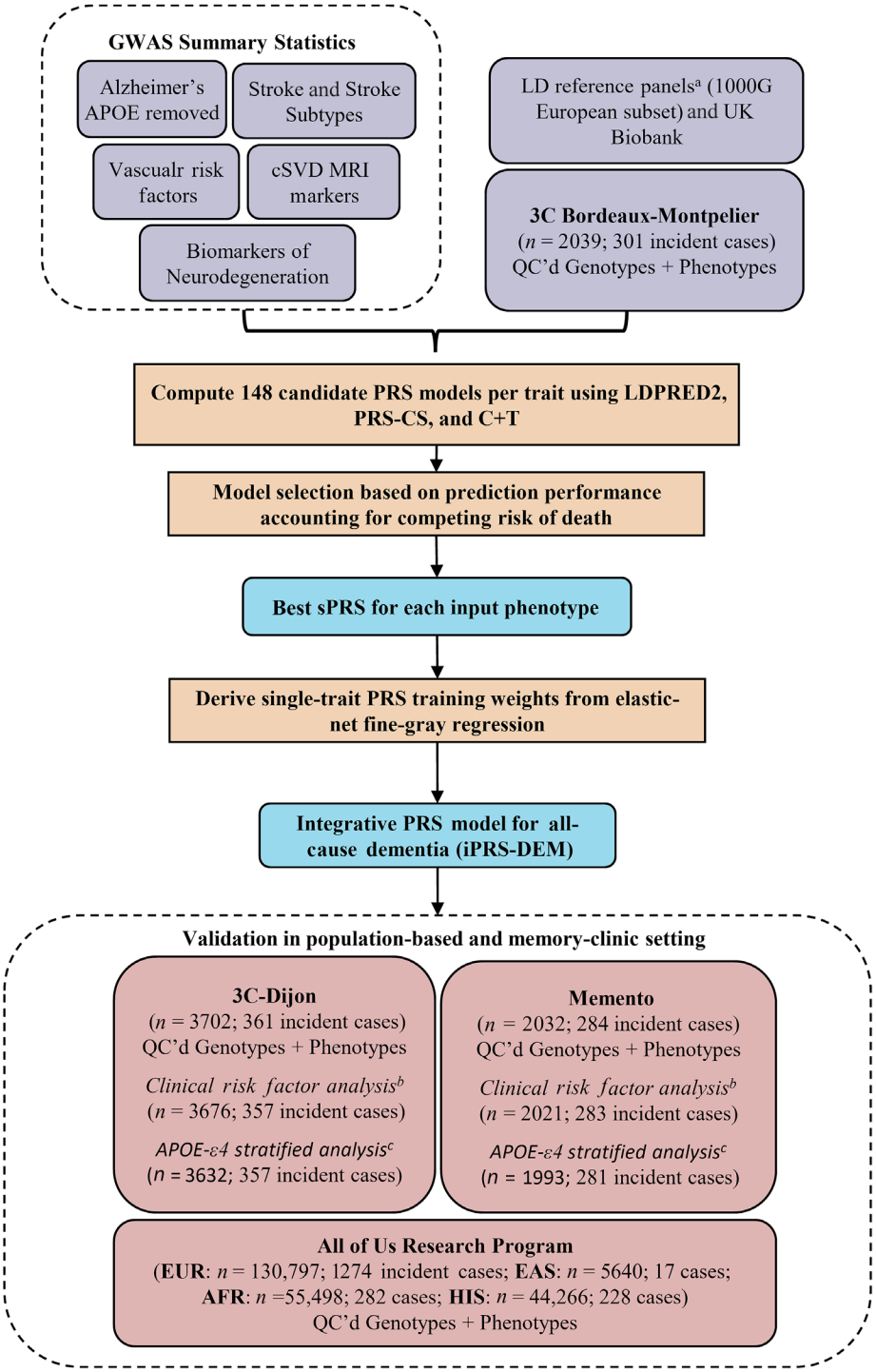


FIGURE 1 Overview of derivation and validation of iPRS-DEM. At each step of iPRS-DEM generation, we used Fine-Gray regression models. LDPRED2, PRS-CS, and C+T are validated methods to derive PRS using distinct combinations of tuning parameters. Candidate PRS models for each trait were selected based on integrative Brier score across 5-, 7.5-, and 10-year horizons. For each selected single-trait PRS the penalized coefficients from elastic-net were divided by the empirical standard deviation to derive training weights. Training weights were used to re-weight the effect size of SNPs included in their respective single-trait PRS. [a]UK Biobank was used as a reference panel for LDPRED2 PRS models as provided by the software. The 1000G European Subset was used as a reference panel for both PRS-CS and C+T. [b]Clinical risk factor analyses included assessing the change in the estimate of iPRS-DEM after adjustment for risk factors and prediction performance at 10-year follow-up in 3C and 5-years in Memento. In Memento, 1,952 participants had complete clinical risk factor data; however, only family history of dementia and low education were used in prediction models (positively associated with dementia risk) corresponding to 2,021 individuals in this analysis. [c]In analyses stratified by *APOE* ε4, we removed those with *APOE* ε2/ε4 genotype. These analyses included interaction models and stratifying individuals into genetic risk groups. AFR, African-American ancestry; C+T, clumping and thresholding; cSVD, cerebral small vessel disease; EAS, East Asian ancestry; EUR, European ancestry; HIS, Hispanic ancestry; LD, linkage disequilibrium; PRS, polygenic risk score(s); QC, quality control; sPRS, single-trait polygenic risk score(s).

15525279, 2025, 3, Downloaded from https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.70014 by Cochrane France, Wiley Online Library on [21/10/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

we estimated the association of iPRS-DEM (per 1 SD) with cumulative incidence of dementia across European (EUR), East Asian (EAS), African-American (AFR), and Hispanic (HIS) ancestries, adjusting for baseline variables (age, sex, 5 PCs, and *APOE* ε4 and ε2 dosage).

3 | RESULTS

3.1 | Description of study populations

After excluding participants without viable genotyping, with prevalent dementia at baseline, or no follow-up (eFigure 1), study populations comprised 5741 3C participants, 236,201 in AoU, and 2032 Memento participants. 3C-BM (N = 2039; mean age ± SD = 73.9 ± 5.1 years; 60% women) had 301 (14.8%) incident dementia cases over median (interquartile range [IQR]) years of follow-up of 9.77 (IQR: 5.7 to 11.6), while 3C-Dijon (N = 3702; 74.2 ± 5.5 years; 61.8% women) had 361 (9.8%) incident dementia cases over 8.4 (IQR: 4.0 to 10.7) years follow-up. 3C-Dijon participants were slightly more educated than 3C-BM participants, but other baseline characteristics, including clinical risk factors and *APOE* genotype, were similar (Table 1). Memento (N = 2032; 70.9 ± 8.6 years; 61.5% women) had 284 (14.0%) dementia cases over 5.0 (IQR: 3.0 to 5.1) years follow-up and had a wider age range, higher education levels, and proportionally more *APOE* ε4 carriers than 3C (Table 1). Per ancestry, AoU had a sample size of 130,797 Europeans (1274 incident dementia cases; 56 ± 17 years, 59.8% women), 5640 East Asians (17 incident dementia cases; 44 ± 17 years; 63.7% women), 55,498 African-Americans (282 incident dementia cases; 49 ± 15 years; 57.6% women), and 44,266 Hispanics (228 incident dementia cases; 45 ± 16 years; 66.8% women). AoU had a larger age range than 3C-Dijon or Memento, and follow-up time was shorter (median of 3 years in each ancestry group, Table 2). In AoU Europeans, the *APOE* genotype distribution was similar to 3C-Dijon, while it was markedly different in other ancestries (*APOE* ε4 being less common in EAS and more common in AFR participants, Table 2).

3.2 | Construction of iPRS-DEM

We selected the best PRS for 27 dementia-related traits (AD, stroke, vascular risk factors, MRI markers of cSVD, and biomarkers of neurodegeneration), based on their ability to predict dementia risk in 3C-BM, and removed three traits whose best PRS did not improve prediction beyond *APOE* (eTable 3). Expectedly, there was considerable correlation between some of the sPRS in the training dataset (eFigure 2). Following elastic-net regression, 12 of 24 sPRS, including a substantial vascular component, had a non-zero training weight and were included in the final iPRS-DEM (eTable 4, eFigure 3, weights of top 250 SNPs are available in eTable 5).

3.3 | Association of iPRS-DEM with cumulative incidence of all-cause dementia in 3C-Dijon

In 3C-Dijon, a 1 SD increase in iPRS-DEM was associated with dementia risk independent of *APOE* ε4 and ε2 dosage (sHR_{per1SD} = 1.15, 95% CI: 1.03 to 1.28; p = .01), and participants in the highest iPRS-DEM quintile had significantly higher dementia risk relative to the middle 20% (sHR = 1.48, 95% CI: 1.08 to 2.05; p = .016). The AD-PRS was not significantly associated with dementia risk for both per 1 SD increase (sHR_{per1SD} = 1.11, 95% CI: 1.00 to 1.23; p = .06) and comparing the upper to middle quintiles (sHR = 1.21, 95% CI: 0.88 to 1.67; p = .25). Sensitivity analysis in cause-specific Cox models revealed similar associations between iPRS-DEM and dementia incidence (eTable 6). The sHR increased when dichotomizing iPRS-DEM at increasing percentiles, reaching sHR = 1.36 (95% CI: 1.09 to 1.68) in the top 30%, sHR = 1.53 (95% CI: 1.04 to 2.26) in the top 5% and sHR = 1.82 (95% CI: 0.92 to 3.62) in the top 1% compared to the rest of the sample (Figure 2A), while no such trend was observed for AD-PRS, with sHRs of 1.37 (95% CI: 1.11 to 1.70), 1.32 (95% CI: 0.86 to 2.01), and 0.77 (95% CI: 0.24 to 2.42) for AD-PRS in the top 30%, 5%, and 1% (eFigure 4). The iPRS-DEM sHR per 1 SD was comparable to that of associated clinical risk factors (sHR_{per1SD} range: 1.14 to 1.24, eTable 7). Adjusting for education and vascular risk factors only modestly attenuated the association of iPRS-DEM with cumulative incidence of dementia (by 27%, eTable 8).

In sex-stratified analyses, we found iPRS-DEM tended to be associated more strongly with dementia risk in women than in men, though interaction was non-significant (iPRS per 1 SD sHR_{Women} = 1.23, 95% CI: 1.07 to 1.41; sHR_{Men} = 1.04, 95% CI: 0.88 to 1.23; interaction p = .13) (Figure 3, eTable 9). iPRS-DEM showed no difference in association with earlier- versus later-onset dementia (< vs ≥ 80 years) (Figure 3, eTable 9).

Next, we explored the risk stratification ability of iPRS-DEM alongside *APOE* ε4. Across *APOE* ε4 carriers and non-carriers there were similar sHRs of iPRS-DEM per 1 SD increase with dementia risk in 3C-Dijon ($sHR_{APOE\,\varepsilon4-}$ = 1.130, 95% CI: 0.99 to 1.28; $sHR_{APOE\,\varepsilon4+}$ = 1.194, 95% CI: 0.98 to 1.46; interaction p = .648). We then stratified individuals into four genetic risk groups by *APOE* ε4 carrier status and iPRS-DEM+/iPRS-DEM−. The optimal cutoff to dichotomize iPRS-DEM was at the 56th percentile in *APOE* ε4+ (sHR = 1.71, 95% CI: 1.11 to 2.62) and at the 94th percentile in *APOE* ε4− (sHR = 1.95, 95% CI: 1.33 to 2.86). In non-parametric estimation of cumulative incidence, iPRS-DEM+ had a significantly higher cumulative incidence of dementia than iPRS-DEM−, in both *APOE* ε4 carriers and non-carriers by age 85 (eTable 10a-c). Overall, cumulative incidence increased with age and greater burden of genetic risk; by age 90 cumulative incidence estimates reached 20.8% (18.0% to 23.6), 24.18% (15.8% to 32.6%), 37.3% (25.6% to 48.7%), and 40.1% (30.7% to 49.4%) in *APOE* ε4-/iPRS-DEM-, *APOE* ε4+/iPRS-DEM-, *APOE* ε4+/iPRS-DEM+, and *APOE* ε4+/iPRS-DEM+, respectively (Figure 2B1, eTable 10a-b). In FGR models, compared to *APOE* ε4-/iPRS-DEM-, both *APOE* ε4+/iPRS-DEM+

15525279, 2025, 3, Downloaded from https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.70014 by Cochrane France, Wiley Online Library on [21/10/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Baseline characteristics of 3C-Bordeaux-Montpellier, 3C-Dijon, and Memento.

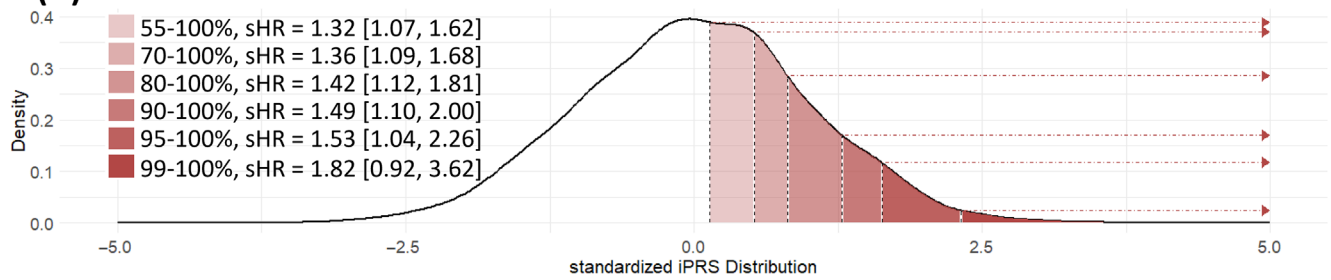
Baseline variables	3C BM (*N* = 2039)	3C Dijon (*N* = 3702)	Memento (*N* = 2032)
Follow-up time (years) Median [IQR]	9.77 [5.66 to 11.6]	8.36 [3.96 to 10.7]	5.00 [3.03 to 5.09]
Status at end of follow-up			
Censored	1427 (70.0%)	2910 (78.6%)	1700 (83.7%)
Demented	301 (14.8%)	361 (9.8%)	284 (14.0%)
Deceased (w/o dementia)	311 (15.3%)	431 (11.6%)	48 (2.4%)
Age at baseline assessment			
Mean (SD)	73.9 (5.09)	74.2 (5.50)	70.9 (8.62)
Median [min, max]	73.3 [65.0 to 92.2]	73.4 [65.0 to 94.6]	71.7 [32.5 to 92.7]
Sex			
Male	816 (40.0%)	1414 (38.2%)	782 (38.5%)
Female	1223 (60.0%)	2288 (61.8%)	1250 (61.5%)
APOE genotype			
ε2/ε2	9 (0.4%)	24 (0.6%)	9 (0.4%)
ε2/ε3	249 (12.2%)	444 (12.0%)	200 (9.8%)
ε2/ε4	21 (1.0%)	70 (1.9%)	39 (1.9%)
ε3/ε3	1383 (67.8%)	2460 (66.5%)	1218 (59.9%)
ε3/ε4	364 (17.9%)	664 (17.9%)	495 (24.4%)
ε4/ε4	13 (0.6%)	40 (1.1%)	71 (3.5%)
Low education (primary education only or no formal)	643 (31.5%)	761 (20.6%)	242 (11.9%)
Missing	5 (0.2%)	1 (0.0%)	0
Hypertension	1559 (76.5%)	2946 (79.6%)	1221 (60.1%)
Missing	0	0	39 (1.9%)
Smoking status			
Never smoker	1921 (94.2%)	3513 (94.9%)	1892 (93.1%)
Current smoker	117 (5.7%)	188 (5.1%)	139 (6.8%)
Missing	1 (0.0%)	1 (0.0%)	1 (0.0%)
Obesity (> 30 kg/m^2)	296 (14.5%)	497 (13.4%)	265 (13.0%)
Missing	23 (1.1%)	9 (0.2%)	36 (1.8%)
Diabetes	209 (10.3%)	332 (9.0%)	177 (8.7%)
Missing	11 (0.5%)	9 (0.2%)	8 (0.4%)
History of CVD	201 (9.9%)	342 (9.2%)	380 (18.7%)
Missing	0 (0%)	2 (0.1%)	8 (0.4%)
Dyslipidemia	1140 (55.9%)	2174 (58.7%)	581 (28.6%)
Missing	2 (0.1%)	5 (0.1%)	0
Family history of dementia[a]	N/A	N/A	850 (41.8%)
Missing	N/A	N/A	11 (0.5%)

Note: Definitions of clinical risk factors differ slightly between cohorts, see eMethods 4 and 10 for details.
Abbreviations: 3C, Three-Cities study; BM, Bordeaux and Montpellier; CVD, cardiovascular disease; IQR, interquartile range.
[a]Family history of dementia was not ascertainable in 3C.

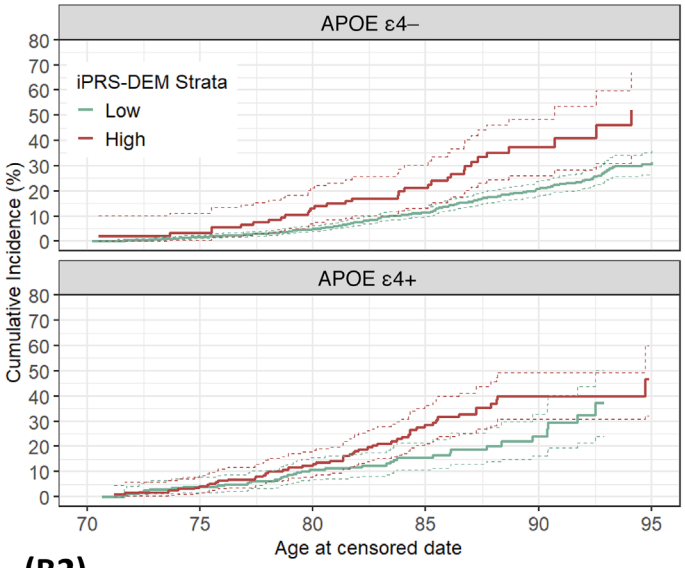
15525279, 2025, 3, Downloaded from https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.70014 by Cochrane France, Wiley Online Library on [21/10/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

15525279, 2025, 3, Downloaded from https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.70014 by Cochrane France, Wiley Online Library on [21/10/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(A)



(B1)



Genetic risk group	Age in years	Estimated cumulative incidence	Number at risk
APOE ε4– and Low PRS (< 93.8th percentile)	75	1.7% [0.9 to 2.4]	1359
	80	4.9% [3.8 to 6.1]	1247
	85	11.5% [9.7 to 13.3]	601
	90	20.8% [18 to 23.6]	199
APOE ε4– and High PRS (> 93.8th percentile)	75	3.5% [-1.4 to 8.4]	81
	80	13.2% [6 to 20.4]	89
	85	21.1% [12.4 to 29.7]	46
	90	37.3% [25.9 to 48.6]	13
APOE ε4+ and Low PRS (< 55.6th percentile)	75	4.1% [1.1 to 7]	211
	80	10.7% [6.3 to 15]	160
	85	15.5% [10.1 to 20.9]	74
	90	24.2% [15.8 to 32.6]	17
APOE ε4+ and High PRS (> 55.6th percentile)	75	4.3% [0.9 to 7.6]	165
	80	12.4% [7.4 to 17.4]	135
	85	28.5% [21 to 36.1]	53
	90	40.1% [30.7 to 49.4]	10

(B2)

Genetic risk strata	*N*	No. cases	sHR (95% CI)	*P*
APOE ε4– and Low iPRS-DEM	2741	234	Reference	Reference
APOE ε4– and High iPRS-DEM	187	33	1.972 [1.361 to 2.858]	0.00033
APOE ε4+ and Low iPRS-DEM	388	38	1.29 [0.909 to 1.832]	0.15
APOE ε4+ and High iPRS-DEM	316	52	2.233 [1.644 to 3.032]	2.7E-07

(C)

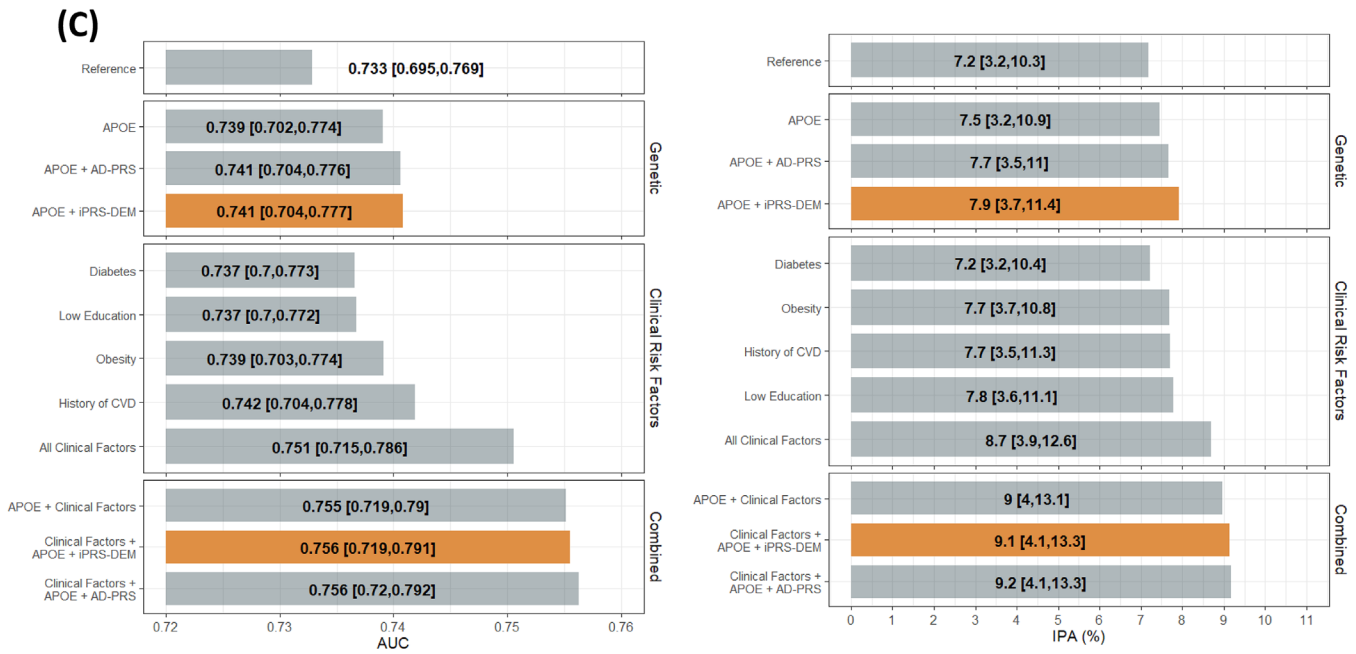


TABLE 2 Per ancestry baseline characteristics of All of Us.

Baseline variables	EAS (*N* = 5640)	AFR (*N* = 55,498)	HIS (*N* = 44,266)	EUR (*N* = 130,797)
Follow-up time (years)*				
Median [IQR]	3.05 (2.04 to 4.05)	3.05 (3.05 to 3.05)	3.05 (2.04 to 3.05)	3.05 (2.04 to 4.05)
Status at end of follow-up				
Censored	5595 (99.2%)	54,760 (98.7%)	43,828 (99.0%)	128,337 (98.1%)
Demented	17 (0.3%)	282 (0.5%)	228 (0.5%)	1274 (1.0%)
Deceased (w/o dementia)	28 (0.5%)	456 (0.8%)	210 (0.5%)	1186 (0.9%)
Age at baseline				
Mean (SD)	44 (17)	49 (15)	45 (16)	56 (17)
Median [IQR]	42 (29 to 58)	52 (38 to 60)	44 (31 to 57)	59 (42 to 69)
Sex				
Male	2050 (36.3%)	23,521 (42.4%)	14,705 (33.2%)	52,559 (40.2)
Female	3590 (63.7%)	31,977 (57.6%)	29,561 (66.8%)	78,238 (59.8%)
APOE Genotype				
ε2/ε2	41 (0.7%)	66 (1.2%)	94 (0.2%)	840 (0.6%)
ε2/ε3	760 (13.5%)	7969 (14.4%)	3161 (7.1%)	16,092 (12.3%)
ε2/ε4	100 (1.8%)	2492 (4.5%)	479 (1.1%)	2791 (2.1%)
ε3/ε3	3827 (67.9%)	25,638 (46.2%)	31,082 (70.2%)	80,765 (61.7%)
ε3/ε4	859 (15.2%)	16,251 (29.2%)	8799 (19.9%)	27,855 (21.4%)
ε4/ε4	53 (0.9%)	2484 (4.5%)	651 (1.5%)	2454 (1.9%)

Abbreviations: AFR, African; EAS, East Asian; EUR, European; HIS, Hispanic.
*Follow-up time is only given on a per-year basis, and not coded as a continuous variable.

and *APOE* ε4−/iPRS-DEM+ conveyed substantially higher dementia risk ($sHR_{APOE\ \varepsilon4+/IPRS\text{-}DEM+}$ = 2.23, 95% CI: 1.64 to 3.03; $sHR_{APOE\ \varepsilon4-/IPRS\text{-}DEM+}$ = 1.97, 95% CI: 1.36 to 2.86]), while the *APOE* ε4+/iPRS-DEM− group was not significantly associated with dementia risk (HR = 1.29, 95% CI: 0.91 to 1.83, Figure 2B2).

3.4 | Predictive performance of iPRS-DEM in 3C-Dijon

At the 10-year follow-up, there was no discernible difference in the prediction of dementia between iPRS-DEM and AD-PRS, both of which performed less well than a model with all clinical factors associated with increased dementia risk in 3C-Dijon (Figure 2C). Combining all clinical risk factors, *APOE ε4* and *ε2* dosage, and iPRS-DEM slightly improved prediction (AUC = 0.756, 95% CI: 0.72 to 0.792; IPA = +9.2% [4.1% to 13.3%]) over a model comprising clinical factors only (AUC = 0.751, 95% CI: 0.715 to 0.786; IPA = +8.7% [3.9 to 12.6%]).

3.5 | External validation of iPRS-DEM in Memento

The iPRS-DEM in Memento included 1,270,010 of 1,320,229 SNPs (96.2%). A 1-SD increase in iPRS-DEM was associated with dementia risk independently of *APOE* ε4 and ε2 dosage (sHR_{per1SD} = 1.25, 95% CI: 1.11 to 1.42; p = .0003, eTable 11). Adjusting for family history of dementia and education, the two clinical risk factors associated with increased all-cause dementia risk in Memento, resulting in negligible

FIGURE 2 Validation of iPRS-DEM in 3C-Dijon. (A) Association of iPRS-DEM with cumulative incidence of dementia across percentile groups. Subdistribution hazard ratios for dementia are per percentile cutoffs relative to the rest of the sample as derived from Fine-Gray regression models. (B) Comparison of genetic risk strata defined by *APOE* ε4 status and iPRS-DEM. (B1) Estimated cumulative incidence curves at ages 75 to 95 years across genetic risk strata. (B2) Association across *APOE* ε4 and iPRS-DEM defined genetic risk strata with cumulative incidence of dementia. All models in each analysis were adjusted for age at baseline (including age-squared), sex, and 10 principal components, as well as *APOE* ε2 and *APOE* ε4 dosage (except for *APOE* ε4 stratified analysis). AD-PRS, Alzheimer's disease polygenic risk score; *APOE*4-, *APOE* ε4 non-carriers; *APOE*4+, *APOE* ε4 carriers; AUC, area under the curve; CVD, cardiovascular disease; IPA, index of prediction accuracy; iPRS, integrative polygenic risk score; sHR, subdistribution hazard ratio. (C) Comparison of prediction performance at 10 years of iPRS-DEM against *APOE*, AD-PRS, and clinical risk factors* based on time-dependent AUC and index of prediction accuracy over 2000 bootstrap replications (*only risk factors showing significant association with increased cumulative incidence of all-cause dementia in 3C-Dijon are used here).

15525279, 2025, 3, Downloaded from https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.70014 by Cochrane France, Wiley Online Library on [21/10/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

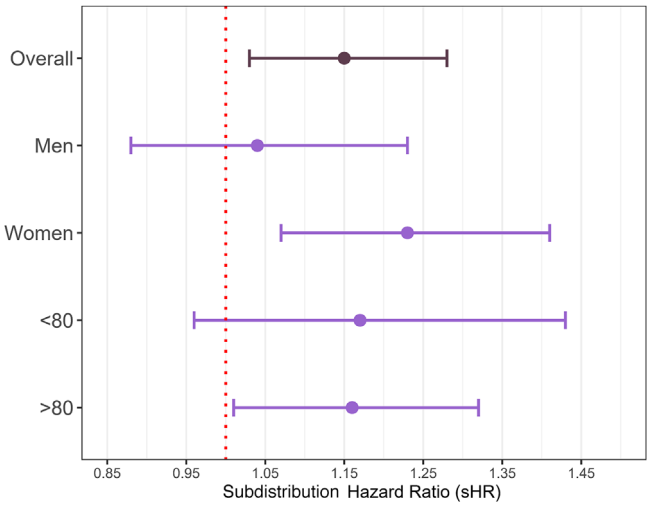


FIGURE 3 iPRS-DEM association with cumulative incidence of dementia in 3C-Dijon, including age and sex subgroups. Error bars represent 95% confidence intervals. All Fine-Gray models are adjusted for age at baseline, sex, 10 genetic principal components, and *APOE* ε2 and *APOE* ε4 dosage. <80 refers to analysis censoring at 80 years old, >80 refers to analysis excluding data before age 80; see eMethods 6 for details.

changes in estimates for the association of iPRS-DEM with dementia risk (4.4%, eTables 12 and 13). Individuals in the highest quintile of iPRS-DEM showed a significant increase in dementia risk relative to the middle 20% (sHR = 1.58, 95% CI: 1.10 to 2.27; $p = .01$). Further, iPRS-DEM showed non-linearly increasing association with dementia risk across higher percentile thresholds, reaching sHRs of 1.79 (95% CI: 1.04 to 2.26), 1.24 (95% CI: 0.74 to 2.09), and 2.28 (95% CI: 0.82 to 6.33) in the top 30%, 5%, and 1% compared to the rest of the sample (Figure 4A). Similar to 3C-Dijon, sex-stratified analyses revealed a non-significant trend toward a stronger association of iPRS-DEM with dementia risk in women than in men (iPRS per 1SD $sHR_{Women} = 1.30$, 95% CI: 1.10 to 1.54; $sHR_{Men} = 1.17$, 95% CI: 0.97 to 1.40; interaction $p = .37$) (Figure 5, eTable 14). Unlike 3C-Dijon, iPRS-DEM showed an association with earlier-onset dementia cases (<80 years) but not in later onset (≥80 years) ($sHR_{<80years} = 1.30$, 95% CI: 1.11 to 1.52; $sHR_{\geq 80years} = 1.13$, 95% CI: 0.92 to 1.38), although, again, confidence intervals overlap (Figure 5, eTable 14).

As in 3C-Dijon, the association of iPRS-DEM per 1 SD increase with dementia risk did not differ significantly between *APOE* ε4 carriers and non-carriers ($sHR_{APOE\,\varepsilon4+} = 1.33$, 95% CI: 1.12 to 1.58; $sHR_{APOE\,\varepsilon4-} = 1.15$, 95% CI: 0.96 to 1.37; $p_{interaction} = .25$). For risk stratification by *APOE* ε4 carrier status and iPRS-DEM+/iPRS-DEM−, the optimal cutoff for iPRS-DEM was at the 69.9th percentile in *APOE* ε4+ ($sHR_{iPRS\text{-}DEM+} = 2.16$, 95% CI: 1.54 to 3.02) and at the 74.1th percentile in *APOE* ε4− ($sHR_{iPRS\text{-}DEM+} = 1.49$, 95% CI: 1.02 to 2.18). Notably, iPRS-DEM+ had a significantly higher cumulative incidence of dementia than iPRS-DEM− by the 3-year follow-up in *APOE* ε4 carriers and by the 4-year follow-up in non-carriers (eTable 15a-c). Estimated cumulative incidence of dementia in each genetic risk strata at the 5-year follow up was 9.8% (7.9% to 11.7%), 13.8% (9.7% to 17.8%), 22.1% (17.8% to 26.4%), and 41.3% (32.9% to 49.8%), in *APOE* ε4-/iPRS-DEM−, *APOE* ε4−/iPRS-DEM+, *APOE* ε4+/iPRS-DEM+, and *APOE* ε4+/iPRS-DEM+, respectively (Figure 4B1). Compared to the *APOE* ε4−/iPRS-DEM− reference group, all genetic risk groups conveyed a significant increase in dementia risk ($sHR_{APOE\,\varepsilon4-/iPRS\text{-}DEM+} = 1.49$, 95% CI: 1.03 to 2.15; $sHR_{APOE\,\varepsilon4+/iPRS\text{-}DEM-} = 2.62$, 95% CI: 1.93 to 3.54; $HR_{APOE\,\varepsilon4+/iPRS\text{-}DEM+} = 5.72$, 95% CI: 4.15 to 7.87, Figure 4B2).

At the 5-year follow-up, iPRS-DEM modestly improved prediction (AUC = 0.758, 95% CI: 0.715 to 0.800; IPA = +10.1% [3.6% to 15.6%]) beyond *APOE* ε4 and ε2 dosage (AUC = 0.751, 95% CI: 0.707 to 0.793; IPA = +9.1% [2.6% to 14.4%]) (Figure 4C). Overall, genetic risk (*APOE* with and without iPRS-DEM) improved dementia prediction above the reference model (age, sex, 10PCs) (AUC = 0.682, 95% CI: 0.636 to 0.727) substantially more than known risk factors, education level and family history, in this setting (AUC = 0.686, 95% CI: 0.641 to 0.731) (Figure 4C).

TABLE 3 Association of iPRS and with dementia within ancestry sub-groups in All of Us.

Ancestry	sHR (95% CI)	*p*
Europeans	1.28 (1.09 to 1.51)	.003
East Asians	5.29 (1.43 to 34.36)	.016
African Americans	1.06 (0.75 to 1.45)	.79
Hispanics	1.09 (0.78 to 1.68)	.48

Fine-Gray models are adjusted for baseline covariates (age, sex, five genetic principal components) and *APOE* dosage.

3.6 | Transportability of iPRS-DEM across ancestries in AoU

The iPRS-DEM in AoU included 1,318,697 of 1,320,229 SNPs (99.9%). 1 SD increase in iPRS-DEM was associated with dementia risk independently of *APOE* ε4 and ε2 dosage in both EUR ($sHR_{per1SD} = 1.28$, [1.09 to 1.51]; $p = .003$) and EAS participants ($sHR_{per1SD} = 5.29$, 95% CI: 1.43 to 34.36; $p = .016$), but not in AFR or HIS, although the direction of effect (+) was consistent (Table 3, Cox model sensitivity in eTable 16).

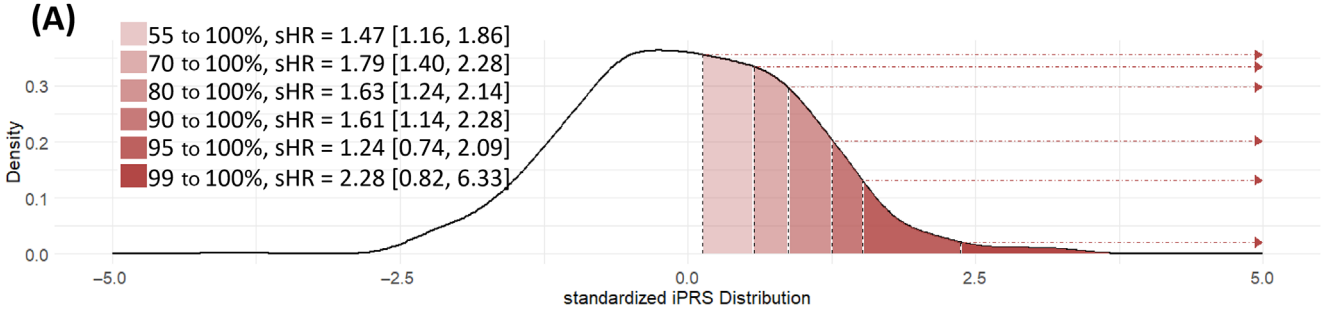
4 | DISCUSSION

We developed a novel integrative PRS for all-cause dementia encompassing genetic risk for multiple traits, in addition to AD, reflecting the neurodegenerative and vascular contribution to dementia risk. We trained iPRS-DEM in an elderly longitudinal population-based cohort and validated it in older community-dwelling people, across ancestries in a large-scale population-based cohort, and in dementia-free memory clinic patients, while accounting for competing risk of death. In older community-dwelling people, iPRS-DEM was associated with dementia risk independently of *APOE* and only modestly attenuated by clinical risk factors, while a single-trait AD-PRS was non-significant. Though improvement in overall prediction was limited, we demonstrated the

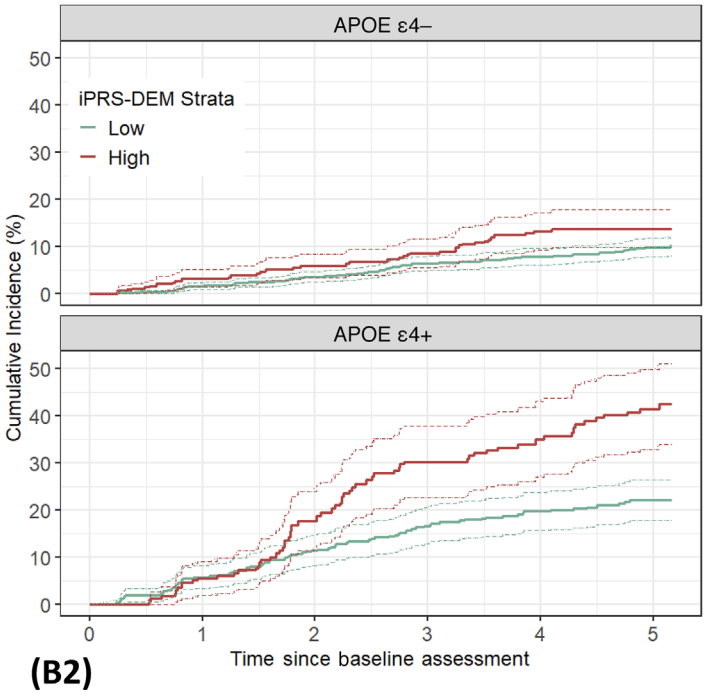
15525279, 2025, 3, Downloaded from https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.70014 by Cochrane France, Wiley Online Library on [21/10/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

15525279, 2025, 3, Downloaded from https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.70014 by Cochrane France, Wiley Online Library on [21/10/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(A)



(B1)

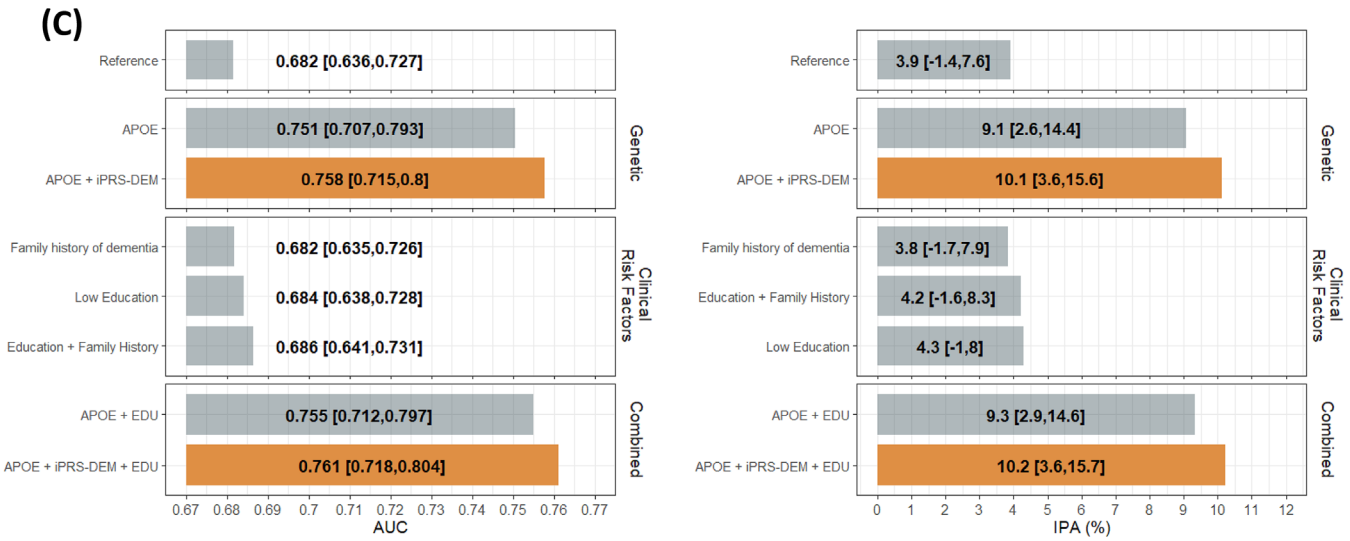


Genetic risk group	Years since baseline	Estimated cumulative incidence	Number at risk
APOE ε4– and Low iPRS-DEM (<74.1th percentile)	1	1.6% [0.8 to 2.4]	1007
	2	3.6% [2.4 to 4.7]	931
	3	6.4% [4.8 to 7.9]	843
	4	7.8% [6.1 to 9.5]	785
	5	9.8% [7.8 to 11.7]	675
APOE ε4– and High iPRS-DEM (>74.1th percentile)	1	3.3% [1.4 to 5.1]	335
	2	5.8% [3.3 to 8.4]	300
	3	8.5% [5.4 to 11.6]	273
	4	13.1% [9.2 to 17]	240
	5	13.8% [9.7 to 17.8]	204
APOE ε4+ and Low iPRS-DEM (<69.9th percentile)	1	5.7% [3.2 to 8.1]	357
	2	11.3% [8.1 to 14.6]	317
	3	16.6% [12.8 to 20.4]	284
	4	19.7% [15.6 to 23.8]	248
	5	22.1% [17.8 to 26.4]	215
APOE ε4+ and High iPRS-DEM (>69.9th percentile)	1	5.5% [1.9 to 9.1]	158
	2	17.5% [11.3 to 23.8]	134
	3	30.1% [22.5 to 37.7]	105
	4	35% [27 to 43]	96
	5	41.3% [32.9 to 49.7]	72

(B2)

Genetic risk strata	*N*	No. cases	sHR (95% CI)	*P*
APOE ε4– and Low iPRS-DEM	1060	91	Reference	Reference
APOE ε4– and High iPRS-DEM	367	45	1.488 [1.028 to 2.154]	0.035
APOE ε4+ and Low iPRS-DEM	393	81	2.616 [1.933 to 3.54]	4.7E-10
APOE ε4+ and High iPRS-DEM	173	64	5.717 [4.151 to 7.874]	1.3E-26

(C)



importance of iPRS-DEM to stratify dementia risk in combination with *APOE* ε4 status. Having a high iPRS-DEM, regardless of *APOE* ε4 status (*APOE* ε4+/iPRS-DEM+ or *APOE* ε4−/iPRS-DEM+) was associated with an approximate two-fold increased risk of dementia compared with *APOE* ε4−/iPRS-DEM−, while *APOE* ε4+/iPRS-DEM− did not convey significantly greater dementia risk. We further validated iPRS-DEM in the memory-clinic setting, showing significant association with dementia risk independent of *APOE* and clinical risk factors and also improving risk prediction and stratification. Particularly in the memory clinic, iPRS-DEM captured substantial genetic risk among *APOE* ε4 carriers, distinguishing cumulative incidence of dementia by the 3-year follow-up, while those with *APOE* ε4+/iPRS-DEM+ had a 5-year cumulative incidence of dementia of 40% and nearly six-fold increased dementia risk compared to the *APOE* ε4−/iPRS-DEM−. We also show strong validation of iPRS-DEM in European adults of all ages and diverse socioeconomic backgrounds. Importantly, we attempted to validate iPRS-DEM across diverse ancestry groups (EAS, AFR, and HIS) and demonstrated transportability to East Asians despite a small sample size. Our findings suggest a promising ability of iPRS-DEM to identify persons at highest genetic risk for dementia in both population and clinical settings.

Polygenic scores may inform dementia prevention strategies by identifying high-risk individuals for either clinical trials or future interventions in the general population.[9,54] However, previous AD-PRS did not explicitly account for the vascular contribution to dementia risk and showed inconsistent relation with all-cause dementia across *APOE* genotypes in population-based cohorts.[19,21,55,56] Our investigations reveal the utility of more comprehensive polygenic risk prediction encompassing genetically determined vascular risk over a single-trait AD-PRS, particularly for all-cause dementia risk stratification in combination with *APOE* ε4.

A recent iPRS for AD did not suggest any benefit over an AD-PRS.[27] However, integrative PRS are sensitive to selection of input traits, methodological choices in PRS derivation, training population, and phenotype definition. The published iPRS focused on AD risk, included a heterogeneous set of traits incompletely capturing vascular risk, and was trained and evaluated in AD case–control datasets that do not consider competing risk of death. In contrast, iPRS-DEM includes input traits specifically focused on neurodegenerative and vascular origins and was trained and evaluated using a longitudinal population-based cohort to predict the probability of all-cause dementia over time using Fine-Gray models to account for competing risk.

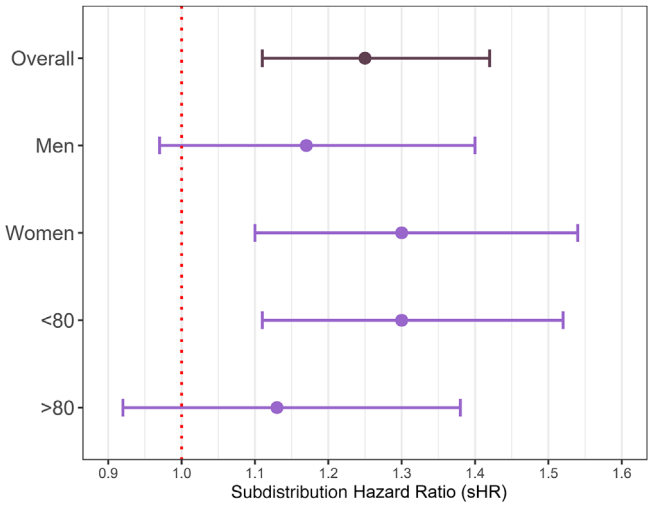


FIGURE 5 iPRS-DEM association with cumulative incidence of dementia in Memento including age and sex subgroups. Error bars represent 95% confidence intervals. All Fine-Gray models are adjusted for age at baseline, sex, 10 genetic principal components, and *APOE ε2* and *APOE ε4* dosage. <80 refers to analysis censoring at 80 years old, >80 refers to analysis excluding data before age 80, see eMethods 6 for details.

Moreover, the previous integrative AD-PRS[27] did not explore risk stratification in combination with *APOE* ε4 that highlighted potential utility of iPRS-DEM in our study.

The limited improvement in prediction performance for all-cause dementia above and beyond clinical factors in a population-based setting is comparable to PRS for other common age-related neurological conditions, like stroke,[24–26] but overall lags behind cardiovascular conditions such as CAD or atrial fibrillation.[11,57] This could be attributed, at least in part, to greater disease heterogeneity. However, iPRS-DEM could potentially still be informative for early preventive interventions at the population level as genetic risk is stable from birth, whereas clinically defined risk factors such as high blood pressure are often detected in mid to late life and may be subject to measurement variability.[58] Further, for population-level risk stratification, we showed the ability of iPRS-DEM to stratify dementia risk in both *APOE* ε4 carriers and non-carriers, with dementia risk in *APOE* ε4 non-carriers with extreme iPRS-DEM being nearly equivalent to that of *APOE* ε4 carriers with high iPRS-DEM.

FIGURE 4 External validation of iPRS-DEM in Memento. (A) Association of iPRS-DEM with cumulative incidence of dementia across percentile groups. Subdistribution hazard ratios for dementia are per percentile cutoffs relative to the rest of the sample as derived from Fine-Gray regression models. Models are adjusted for age at baseline, sex, and 10 principal components, as well as *APOE* ε4 and ε2dosage. (B) Comparison of genetic risk strata defined by *APOE* ε4 status and iPRS-DEM. (B1) Estimated cumulative incidence curves at up to 5-year follow-up across genetic risk strata. (B2) Association across *APOE* ε4 and iPRS-DEM defined genetic risk strata with cumulative incidence of dementia. All models in each analysis are adjusted for age at baseline, sex, and 10 principal components, as well as *APOE* ε2 and *APOE* ε4 dosage (except for *APOE* ε4 stratified analysis). AUC, area under the curve; *APOE*4−, *APOE* ε4 non-carriers; *APOE*4+, *APOE* ε4 carriers; EDU, low education; IPA, index of prediction accuracy; iPRS, integrative polygenic risk score; sHR, subdistribution hazard ratio. (C) Comparison of prediction performance at 5 years of iPRS-DEM against *APOE*, AD-PRS, and clinical risk factors* based on time-dependent area under the curve and index of prediction accuracy over 2000 bootstrap replications (*only risk factors showing significant association with increased cumulative incidence of all-cause dementia in Memento are used here).

15525279, 2025, 3, Downloaded from https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.70014 by Cochrane France, Wiley Online Library on [21/10/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Originally, we extended iPRS-DEM to the memory-clinic setting, exploring its potential to inform dementia risk assessment in persons with MCI or subjective cognitive complaints.[13] Studies investigating PRS in a memory clinic setting are scarce, with one study showing that low single-trait AD-PRS mitigated AD risk in *APOE* ε4 carriers, although it was not associated with all-cause dementia.[59] We provide initial evidence that iPRS-DEM, despite being trained in a population-based setting, is associated with and improves prediction of all-cause dementia in a memory-clinic population. Notably, high iPRS-DEM was associated with dementia risk in both *APOE* ε4 strata and was particularly strong among *APOE* ε4 carriers, distinguishing risk by the 3-year follow-up. Taken together, *APOE* ε4 carriers with high iPRS-DEM approximated a nearly six-fold increased dementia risk compared to *APOE* ε4 non-carriers with "low" iPRS-DEM representing the majority of the population. These results support the idea that iPRS-DEM alongside *APOE* could be a powerful tool to identify memory-clinic patients for targeted preventive interventions or enrollment in clinical trials.

There is mounting, cross-domain evidence for sex differences in AD pathogenesis,[60] which likely extends to polygenic risk.[61] Here, we found in both 3C-Dijon and Memento that iPRS-DEM was associated with dementia risk more strongly (and, significantly, in part due to larger sample size) in women than in men, which is congruent with a recent study investigating AD-PRS in a population-based German cohort (ESTHER).[21] Incidence of AD is known to be higher in women among older adults likely due to greater longevity in women and men having healthier cardiovascular risk profiles at older ages.[62] Given both 3C and Memento have a higher proportion of women (~60%) and comprise largely older individuals, PRS for vascular traits included in iPRS-DEM may, to some extent, capture genetic risk predominantly expressed in women. These findings highlight the need for sex-stratified GWAS of dementia (as well as other relevant traits included in the iPRS) and that future PRS for dementia should consider training models separately in women and men.

Lastly, iPRS-DEM association with dementia risk strongly replicated in European ancestry participants of AoU, importantly spanning a wide age range and with substantial geographical and socioeconomic diversity. We also leveraged the unique genetic diversity in AoU to assess the transportability of iPRS-DEM to non-Europeans, demonstrating a significant association with dementia risk in East Asians. These results did not translate to African American or Hispanic ancestry, although the directionality of effect was consistent. We suggest that different translatability of iPRS-DEM in EAS compared to AFR or HIS ancestry could be driven by increased genetic diversity in AFR[63] and admixture effects in HIS.[64] A recent study in AoU proposed a framework to develop translatable PRS for non-Europeans across multiple conditions, although AD/dementia was not investigated.[65] Further, recent developments in PRS methods that leverage cross-ancestry reference panels (PRSCS-X) and admixture mapping (GAUDI) propose to enhance PRS transportability.[64] Most importantly, our results highlight the crucial need to increase diversity in GWAS used to generate PRS,[26,66] especially for neurodegenerative traits and MRI-based endophenotypes.

We acknowledge the limitations of this study. For certain traits, such as MRI markers of cSVD and other biomarkers, their contribution to iPRS-DEM was likely limited by underpowered GWAS. Of note, 3C was restricted to participants >65 years, stipulating that those at highest risk (ie, high genetic burden) of developing earlier-onset dementia are less likely to be selected for the study. Indeed, *APOE* ε4 homozygosity has been shown to represent a distinct genetic form of AD, with an average age of symptom onset of 65.1 years.[67] Retraining in cohorts spanning wider age ranges would help address this selection issue. Further, we cannot rule out possible bias introduced by interval censoring or selective survival common to dementia studies such as 3C or Memento.[68] We did not investigate the association of iPRS-DEM with dementia subtypes due to limited sample size. Comparison of iPRS-DEM with more comprehensive risk factors for dementia and standard neuropsychological assessment should be considered in future work on larger samples.[7] Ultimately, translation into clinical practice may be optimized by the integration of iPRS-DEM with both clinical variables and other biomarkers, especially novel plasma AD biomarkers,[69] into risk prediction models based on global, multimodal biomarker measurements. Promisingly, CAD PRS have shown benefit alongside clinical risk prediction models to inform statin use in primary prevention.[70]

By accounting for the vascular contribution to dementia risk, iPRS-DEM may provide useful guidance for trial enrichment designs that are more representative of global dementia risk. Our results suggest that in combination with *APOE* ε4, it could offer substantial predictive enrichment by identifying participants at high likelihood of developing dementia.[71] Whether iPRS-DEM could inform prognostic enrichment as well remains unclear. It will also be interesting to examine whether iPRS-DEM can help identify individuals especially predisposed to vascular brain disease who are at high risk of developing amyloid-related imaging abnormalities from amyloid immunotherapy, currently considered for AD prevention in asymptomatic persons with high amyloid burden.[72]

Overall, we provide evidence that an iPRS-DEM incorporating both neurodegenerative and vascular contributions to dementia is a powerful genetic risk stratification tool in addition to *APOE* ε4 that could be applied in both population-based and memory-clinic settings for dementia precision prevention. Future studies refining and validating iPRS-DEM in additional, especially non-European, populations are warranted.

ACKNOWLEDGMENTS

The authors would like to thank all Three-City and Memento participants, as well as the study teams (including the investigators and neurology teams) for their contributions and dedication to the study. They are grateful to the All of Us participants for their contributions; without them, this research would not have been possible. They thank the participating studies of the GIGASTROKE consortium; detailed acknowledgments are provided in the original manuscript. They also are grateful for the provision of summary statistics without the Three-City study, and a list of collaborators is provided in the

15525279, 2025, 3, Downloaded from https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.70014 by Cochrane France, Wiley Online Library on [21/10/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

appendix. Further, the authors would like to thank the ENIGMA consortium for providing summary statistics for hippocampal volume and longitudinal and age-related change of regional brain volumes. They thank the participating studies of the CHARGE consortium; detailed acknowledgments are provided in the original manuscripts in which the publicly available summary statistics were used. This project is supported by a grant overseen by the French National Research Agency (ANR) as part of the Investments for the Future program ANR-18-RHUS-0002 (RHU-SHIVA) and by the Precision and Global Vascular Brain Health Institute (VBHI) funded by the France 2030 IHU3 initiative ANR-23-IAHU-0001. Stéphanie Debette also received funding from the Prix Burrus – Fondation pour la Recherche Médicale and the Prix NRJ-neurosciences Académie des Sciences. Computations were performed on the Bordeaux Bioinformatics Center (CBiB) and CREDIM computer resources, University of Bordeaux. Funding support for additional computing resources was provided to Stéphanie Debette by the Fondation Claude Pompidou. Tim D'Aoust was supported by the Digital Public Health Graduate Program (DPH), a PhD program supported by the French Investments for the Future program (Grant 17-EURE-0019). Support for title page creation and format was provided by AuthorArranger, a tool developed at the National Cancer Institute. Santiago Clocchiatti-Tuozzo is funded by NIH T32 AG019134 and NIH P30AG021342. The MEMENTO cohort is funded by the Fondation Plan Alzheimer (Alzheimer Plan 2008-2012), through the Plan Maladies Neurodégénératives (2014 to 2019), and the French Ministry of Research (MESRI, DGRI 2020-2024). This work was also supported by CIC 1401-EC, Bordeaux University Hospital (CHU Bordeaux, sponsor of the cohort), Inserm, and the University of Bordeaux. Genome-wide genotyping was funded by a grant (EADB) from the EU Joint Programme—Neurodegenerative Disease Research. The Three-City (3C) study is conducted under a partnership agreement among the Institut National de la Santé et de la Recherche Médicale (INSERM), the University of Bordeaux, and Sanofi-Aventis. The Fondation pour la Recherche Médicale funded the preparation and initiation of the study. The 3C Study is also supported by the Caisse Nationale Maladie des Travailleurs Salariés, Direction Générale de la Santé, Mutuelle Générale de l'Education Nationale (MGEN), Institut de la Longévité, Conseils Régionaux of Aquitaine and Bourgogne, Fondation de France, and Ministry of Research–INSERM Programme "Cohortes et collections de données biologiques." The All of Us research program, funded by the National Institutes of Health (NIH) (grant R01HG010001) through the National Center for Advancing Translational Sciences (NCATS) and the National Human Genome Research Institute (NHGRI). This work was funded by a grant (EADB) from the EU Joint Programme—Neurodegenerative Disease Research. Inserm UMR1167 is also funded by the Inserm, Institut Pasteur de Lille, Lille Métropole Communauté Urbaine, and French government's LABEX DISTALZ program (Development of Innovative Strategies for a Transdisciplinary Approach to ALZheimer's disease). A list of collaborators is provided in the appendix. They also are grateful for the provision of summary statistics without the EADI (including the Three-City study) and Memento, and a list of collaborators is provided in the appendix. The ethics committees of the University Hospital of Kremlin-Bicêtre and Dijon University Hospital approved the 3C study protocols, which adhere to the tenets of the Declaration of Helsinki. The Memento study was approved by the ethics committee "Comité de Protection des Personnes Sud-Ouest et Outre Mer III," and was conducted following standards of the Good Clinical Practice and the Helsinki Declaration. All participants of the Three-City and Memento studies provided informed consent at enrollment. The AoU Institutional Review Board approved the protocol and study (https://allofus.nih.gov/about/who-we-are/institutional-review-board-irb-of-all-of-us-research-program), and all participants or their legally designated surrogates provided written informed consent. AoU data are publicly available at www.allofus.nih.gov, and all data access and analyses were conducted within a cybersecure workspace provided by the National Institutes of Health.

CONFLICT OF INTEREST STATEMENT

The authors of this manuscript have no competing interests to declare. Author disclosures are available in the Supporting Information.

ORCID

Tim D'Aoust https://orcid.org/0009-0003-5601-1497

REFERENCES

1. Nichols E, Steinmetz JD, Vollset SE, et al. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study. *Lancet Public Health*. 2019;7(2):e105-e125. doi:10.1016/S2468-2667(21)00249-8
2. Gladman JT, Corriveau RA, Debette S, et al. Vascular contributions to cognitive impairment and dementia: research consortia that focus on etiology and treatable targets to lessen the burden of dementia worldwide. *Alzheimers Dement Transl Res Clin Interv*. 2019;5:789-796. doi:10.1016/j.trci.2019.09.017
3. Snyder HM, Corriveau RA, Craft S, et al. Vascular contributions to cognitive impairment and dementia including Alzheimer's disease. *Alzheimers Dement*. 2015;11(6):710-717. doi:10.1016/j.jalz.2014.10.008
4. Schneider JA, Arvanitakis Z, Bang W, Bennett DA. Mixed brain pathologies account for most dementia cases in community-dwelling older persons. *Neurology*. 2007;69(24):2197-2204. doi:10.1212/01.wnl.0000271090.28148.24
5. van der Flier WM, Skoog I, Schneider JA, et al. Vascular cognitive impairment. *Nat Rev Dis Primer*. 2018;4(1):1-16. doi:10.1038/nrdp.2018.3
6. Kim CK, Lee YR, Ong L, Gold M, Kalali A, Sarkar J. Alzheimer's disease: key insights from two decades of clinical trial failures. *J Alzheimers Dis*. 2022;87(1):83-100. doi:10.3233/JAD-215699
7. Livingston G, Huntley J, Sommerlad A, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet. *Lancet*. 2020;396(10248):413-446. doi:10.1016/S0140-6736(20)30367-6
8. Solomon A, Turunen H, Ngandu T, et al. Effect of the Apolipoprotein E genotype on cognitive change during a multidomain lifestyle intervention: a subgroup analysis of a randomized clinical trial. *JAMA Neurol*. 2018;75(4):462. doi:10.1001/jamaneurol.2017.4365
9. Kivipelto M, Mangialasche F, Ngandu T. Lifestyle interventions to prevent cognitive impairment, dementia and Alzheimer disease. *Nat Rev Neurol*. 2018;14(11):653-666. doi:10.1038/s41582-018-0070-3
10. Frisoni GB, Molinuevo JL, Altomare D, et al. Precision prevention of Alzheimer's and other dementias: anticipating future needs in the control of risk factors and implementation of disease-modifying therapies. *Alzheimers Dement*. 2020;16(10):1457-1468. doi:10.1002/alz.12132

15525279, 2025, 3, Downloaded from https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.70014 by Cochrane France, Wiley Online Library on [21/10/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

11. Khera AV, Chaffin M, Aragam KG, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet*. 2018;50(9):1219-1224. doi:10.1038/s41588-018-0183-z
12. Cholerton B, Larson EB, Quinn JF, et al. Precision medicine: clarity for the complexity of dementia. *Am J Pathol*. 2016;186(3):500-506. doi:10.1016/j.ajpath.2015.12.001
13. Frisoni GB, Altomare D, Ribaldi F, et al. Dementia prevention in memory clinics: recommendations from the European task force for brain health services. *Lancet Reg Health—Eur*. 2023;26:100576. doi:10.1016/j.lanepe.2022.100576
14. Fahed AC, Philippakis AA, Khera AV. The potential of polygenic scores to improve cost and efficiency of clinical trials. *Nat Commun*. 2022;13(1):2922. doi:10.1038/s41467-022-30675-z
15. Harrison JR, Mistry S, Muskett N, Escott-Price V. From polygenic scores to precision medicine in Alzheimer's disease: a systematic review. *J Alzheimers Dis*. 2020;74(4):1271-1283. doi:10.3233/JAD-191233
16. Stocker H, Möllers T, Perna L, Brenner H. The genetic risk of Alzheimer's disease beyond APOE ε4: systematic review of Alzheimer's genetic risk scores. *Transl Psychiatry*. 2018;8(1):1-9. doi:10.1038/s41398-018-0221-8
17. Ware EB, Faul JD, Mitchell CM, Bakulski KM. Considering the APOE locus in Alzheimer's disease polygenic scores in the Health and Retirement Study: a longitudinal panel study. *BMC Med Genomics*. 2020;13(1):164. doi:10.1186/s12920-020-00815-9
18. Stocker H, Perna L, Weigl K, et al. Prediction of clinical diagnosis of Alzheimer's disease, vascular, mixed, and all-cause dementia by a polygenic risk score and APOE status in a community-based cohort prospectively followed over 17 years. *Mol Psychiatry*. 2021;26(10):5812-5822. doi:10.1038/s41380-020-0764-y
19. Najar J, van der Lee SJ, Joas E, et al. Polygenic risk scores for Alzheimer's disease are related to dementia risk in APOE ε4 negatives. *Alzheimers Dement Diagn Assess Dis Monit*. 2021;13(1):e12142. doi:10.1002/dad2.12142
20. Riaz M, Huq A, Ryan J, et al. Effect of APOE and a polygenic risk score on incident dementias and cognitive decline in a healthy older population. *Aging Cell*. 2021;20(6):e13384. doi:10.1111/acel.13384
21. Stocker H, Trares K, Beyer L, et al. Alzheimer's disease risk, and dementia-related blood biomarker levels in a population-based cohort study followed over 17 years. *Alzheimers Res Ther*. 2023;15(1):129. doi:10.1186/s13195-023-01277-8. Alzheimer's polygenic risk scores, APOE.
22. Ikram MA, Bersano A, Manso-Calderón R, et al. Genetics of vascular dementia – review from the ICVD working group. *BMC Med*. 2017;15(1):48. doi:10.1186/s12916-017-0813-9
23. Inouye M, Abraham G, Nelson CP, et al. Genomic risk prediction of coronary artery disease in 480,000 adults. *J Am Coll Cardiol*. 2018;72(16):1883-1893. doi:10.1016/j.jacc.2018.07.079
24. Abraham G, Malik R, Yonova-Doing E, et al. Genomic risk score offers predictive performance comparable to clinical risk factors for ischaemic stroke. *Nat Commun*. 2019;10(1):5819. doi:10.1038/s41467-019-13848-1
25. Myserlis EP, Georgakis MK, Demel SL, et al. A genomic risk score identifies individuals at high risk for intracerebral hemorrhage. *Stroke*. 2023;54(4):973-982. doi:10.1161/STROKEAHA.122.041701
26. Mishra A, Malik R, Hachiya T, et al. Stroke genetics informs drug discovery and risk prediction across ancestries. *Nature*. 2022;611(7934):115-123. doi:10.1038/s41586-022-05165-3
27. Clark K, Fu W, Liu CL, et al. The prediction of Alzheimer's disease through multi-trait genetic modeling. *Front Aging Neurosci*. 2023;15:1168638. doi:10.3389/fnagi.2023.1168638
28. Vascular factors and risk of dementia: design of the three-city study and baseline characteristics of the study population. *Neuroepidemiology*. 2003;22(6):316-325. doi:10.1159/000072920
29. the MEMENTO cohort Study Group, Dufouil C, Dubois B, et al. Cognitive and imaging markers in non-demented subjects attending a memory clinic: study design and baseline findings of the MEMENTO cohort. *Alzheimers Res Ther*. 2017;9(1):67. doi:10.1186/s13195-017-0288-0
30. The All of Us Research Program Investigators. All of us research program. *N Engl J Med*. 2019;381(7):668-676. doi:10.1056/NEJMsr1809937
31. Ramirez AH, Sulieman L, Schlueter DJ, et al. The *All of Us* Research Program. Data quality. *Patterns*. 2022;3(8):100570. doi:10.1016/j.patter.2022.100570
32. NLoM. Accessed January 15, 2024. Overview of Systematized Nomenclature of Medicine (SNOMED). https://www.nlm.nih.gov/healthit/snomedct/snomed_overview.html
33. World Health Organization. *International Statistical Classification of Diseases and Related Health Problems*. 10th ed. World Health Organization; 2014. https://icd.who.int/browse10/2016/en
34. Choi SW, Mak TSH, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. *Nat Protoc*. 2020;15(9):2759-2772. doi:10.1038/s41596-020-0353-1
35. Leonenko G, Baker E, Stevenson-Hoare J, et al. Identifying individuals with high risk of Alzheimer's disease using polygenic risk scores. *Nat Commun*. 2021;12(1):4506. doi:10.1038/s41467-021-24082-z
36. Choi SW, O'Reilly PF. PRSice-2: polygenic risk score software for biobank-scale data. *GigaScience*. 2019;8(7):giz082. doi:10.1093/gigascience/giz082
37. Privé F, Arbel J, Vilhjálmsson BJ. LDpred2: better, faster, stronger. *Bioinformatics*. 2021;36(22-23):5424-5431. doi:10.1093/bioinformatics/btaa1029
38. Ge T, Chen CY, Ni Y, Feng YCA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun*. 2019;10(1):1776. doi:10.1038/s41467-019-09718-5
39. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc*. 1999;94(446):496-509. doi:10.1080/01621459.1999.10474144
40. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006;38(8):904-909. doi:10.1038/ng1847
41. Park SY, Park JE, Kim H, Park SH. Review of statistical methods for evaluating the performance of survival or other time-to-event prediction models (from conventional to deep learning approaches). *Korean J Radiol*. 2021;22(10):1697-1707. doi:10.3348/kjr.2021.0223
42. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc Ser B Stat Methodol*. 2005;67(2):301-320. doi:10.1111/j.1467-9868.2005.00503.x
43. Simon N, Friedman JH, Hastie T, Tibshirani R. Regularization paths for Cox's proportional hazards model via coordinate descent. *J Stat Softw*. 2011;39:1-13. doi:10.18637/jss.v039.i05
44. Friedman JH, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33:1-22. doi:10.18637/jss.v033.i01
45. Austin PC, Lee DS, Fine JP. Introduction to the analysis of survival data in the presence of competing risks. *Circulation*. 2016;133(6):601-609. doi:10.1161/CIRCULATIONAHA.115.017719
46. Austin PC, Fine JP. Practical recommendations for reporting Fine-Gray model analyses for competing risk data. *Stat Med*. 2017;36(27):4391-4400. doi:10.1002/sim.7501
47. Meister R, Schaefer C. Statistical methods for estimating the probability of spontaneous abortion in observational studies—Analyzing pregnancies exposed to coumarin derivatives. *Reprod Toxicol*. 2008;26(1):31-35. doi:10.1016/j.reprotox.2008.06.006
48. Allignol A, Schumacher M, Beyersmann J. Empirical transition matrix of multi-state models: the etm package. *J Stat Softw*. 2011;38:1-15. doi:10.18637/jss.v038.i04

15525279, 2025, 3, Downloaded from https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.70014 by Cochrane France, Wiley Online Library on [21/10/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

49. Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat Med*. 2013;32(30):5381-5397. doi:10.1002/sim.5958
50. Kattan MW, Gerds TA. The index of prediction accuracy: an intuitive measure useful for evaluating risk prediction models. *Diagn Progn Res*. 2018;2(1):7. doi:10.1186/s41512-018-0029-2
51. Ozenne B, Sørensen AL, Scheike T, Torp-Pedersen C, Gerds TA. riskRegression: predicting the risk of an event using Cox regression models. *R J*. 2017;9(2):440-460.
52. Denz R, Klaaßen-Mielke R, Timmesfeld N. A comparison of different methods to adjust survival curves for confounders. *Stat Med*. 2023;42(10):1461-1479. doi:10.1002/sim.9681
53. Core Team R. *R: A Language and Environment for Statistical Computing; 2021*. https://www.R-project.org/
54. Rasmussen J I, Frikke-Schmidt R. Modifiable cardiovascular risk factors and genetics for targeted prevention of dementia. *Eur Heart J*. 2023;44(28):2526-2543. doi:10.1093/eurheartj/ehad293
55. Yu C, Ryan J, Orchard SG, et al. Validation of newly derived polygenic risk scores for dementia in a prospective study of older individuals. *Alzheimers Dement*. 2023;19(12):5333-5342. doi:10.1002/alz.13113
56. der LeeSJvan, FJ Wolters, Ikram MK, et al. The effect of APOE and other common genetic variants on the onset of Alzheimer's disease and dementia: a community-based cohort study. *Lancet Neurol*. 2018;17(5):434-444. doi:10.1016/S1474-4422(18)30053-X
57. O'Sullivan JW, Ashley EA, Elliott PM. Polygenic risk scores for the prediction of cardiometabolic disease. *Eur Heart J*. 2023;44(2):89-99. doi:10.1093/eurheartj/ehac648
58. Debette S, Seshadri S. Vascular risk factors and dementia revisited. *J Neurol Neurosurg Psychiatry*. 2009;80(11):1183-1184. doi:10.1136/jnnp.2009.181289
59. Ebenau JL, van der Lee SJ, Hulsman M, et al. Risk of dementia in APOE ε4 carriers is mitigated by a polygenic risk score. *Alzheimers Dement Diagn Assess Dis Monit*. 2021;13(1):e12229. doi:10.1002/dad2.12229
60. Kolahchi Z, Henkel N, Eladawi MA, et al. Sex and gender differences in Alzheimer's disease: genetic, hormonal, and inflammation impacts. *Int J Mol Sci*. 2024;25(15):8485. doi:10.3390/ijms25158485
61. Fan CC, Banks SJ, Thompson WK, et al. Sex-dependent autosomal effects on clinical progression of Alzheimer's disease. *Brain*. 2020;143(7):2272-2280. doi:10.1093/brain/awaa164
62. Chêne G, Beiser A, Au R, et al. Gender and incidence of dementia in the Framingham Heart Study from mid-adult life. *Alzheimers Dement*. 2015;11(3):310-320. doi:10.1016/j.jalz.2013.10.005
63. Majara L, Kalungi A, Koen N, et al. Low and differential polygenic score generalizability among African populations due largely to genetic diversity. *Hum Genet Genomics Adv*. 2023;4(2):100184. doi:10.1016/j.xhgg.2023.100184
64. Wang Y, Tsuo K, Kanai M, Neale BM, Martin AR. Challenges and opportunities for developing more generalizable polygenic risk scores. *Annu Rev Biomed Data Sci*. 2022;5:293-320. doi:10.1146/annurev-biodatasci-111721-074830
65. Lennon NJ, Kottyan LC, Kachulis C, et al. Selection, optimization and validation of ten chronic disease polygenic risk scores for clinical implementation in diverse US populations. *Nat Med*. 2024;30(2):480-487. doi:10.1038/s41591-024-02796-z
66. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet*. 2019;51(4):584-591. doi:10.1038/s41588-019-0379-x
67. Fortea J, Pegueroles J, Alcolea D, et al. APOE4 homozygozity represents a distinct genetic form of Alzheimer's disease. *Nat Med*. 2024;6:1-8. doi:10.1038/s41591-024-02931-w. Published online May.
68. Weuve J, Proust-Lima C, Power MC, et al. Guidelines for reporting methodological challenges and evaluating potential bias in dementia research. *Alzheimers Dement*. 2015;11(9):1098-1109. doi:10.1016/j.jalz.2015.06.1885
69. Hansson O, Blennow K, Zetterberg H, Dage J. Blood biomarkers for Alzheimer's disease in clinical practice and trials. *Nat Aging*. 2023;3(5):506-519. doi:10.1038/s43587-023-00403-3
70. Marston NA, Pirruccello JP, Melloni GEM, et al. Predictive utility of a coronary artery disease polygenic risk score in primary prevention. *JAMA Cardiol*. 2023;8(2):130-137. doi:10.1001/jamacardio.2022.4466
71. Ballard C, Atri A, Boneva N, et al. Enrichment factors for clinical trials in mild-to-moderate Alzheimer's disease. *Alzheimers Dement Transl Res Clin Interv*. 2019;5:164-174. doi:10.1016/j.trci.2019.04.001
72. Walsh T, Duff L, Riviere ME, et al. Outreach, screening, and randomization of APOE ε4 carriers into an Alzheimer's prevention trial: a global perspective from the API Generation Program. *J Prev Alzheimers Dis*. 2023;10(3):453-463. doi:10.14283/jpad.2023.27

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: D'Aoust T, Clocchiatti-Tuozzo S, Rivier CA, et al. Polygenic score integrating neurodegenerative and vascular risk informs dementia risk stratification. *Alzheimer's Dement*. 2025;21:e70014. https://doi.org/10.1002/alz.70014

15525279, 2025, 3, Downloaded from https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.70014 by Cochrane France, Wiley Online Library on [21/10/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Chapter 5 – PRS to inform secondary stroke prevention

In the second article, we move from a primary prevention setting for dementia to a secondary stroke prevention setting by investigating PRS ability to predict adverse outcomes in the PRESTIGE-AF clinical trial. PRESTIGE-AF trial is a phase-3 RCT aiming to assess the efficacy and safety of DOAC in survivors of ICH that also have AF (102). This trial was funded by the European Union's Horizon 2020 research and innovation programme. Pr Stéphanie Debette was responsible for the genetic sub-study. As part of my PhD, I joined two investigator meetings at Imperial College of London where I presented first an analysis plan in November 2023 and then initial results of the sub-study in November 2024. I was responsible for all analyses, including performing quality control and imputation of the genotyped data provided by our colleagues at the Clinical Research Laboratory and Biobank – Genetic and Molecular Epidemiology Laboratory at McMaster University in Canada. This work was presented at the European Stroke Organization Conference in May 2025 in Helsinki, Finland. The article presented here aligns with deliverables outlined at the outset of the PRESTIGE-AF trial.

5.1 Polygenic risk informs anticoagulation treatment in intracerebral hemorrhage survivors with atrial fibrillation

DOACs are the first-line therapy for ischemic stroke prevention for patients with AF. However, optimal stroke prevention strategies remain uncertain for AF patients with high bleeding risk, such as ICH survivors. Approximately 20-25% of ICH survivors have AF and are often anticoagulated at the time of ICH. Essentially, we are faced with a clinical dilemma as these patients face simultaneous high risk of bleeding and of ischemic stroke due to AF, and lack clear guidance on resuming (or starting) anticoagulation. Furthermore, ICH survivors with AF have been mostly excluded from large trials of stroke prevention in AF individuals. The PRESTIGE-AF trial (NCT03996772) became the first completed, phase 3 randomized clinical trial (N=319) aiming to assess if DOAC is effective at preventing ischemic stroke and an equally safe option with respect to recurrent ICH. However, the trial showed DOACs reduced ischemic stroke risk but this benefit was offset by increased major hemorrhagic complications, including recurrent ICH, without establishing non-inferiority for ICH risk (102). A recent meta-analysis including PRESTIGE-AF and two of the earlier stage clinical trials, SoSTART and APACHE-AF, showed a significant increase in recurrent ICH and major bleeding without a significant reduction in stroke or systemic

embolism (103). These mixed findings underscore the challenge of balancing ischemic and bleeding risks in this complex patient population. PRS for stroke and associated traits may improve stroke risk prediction and stratification in clinical trials (10,22), including among those with existing clinical indications (213,215). Importantly, PRS also has potential to inform treatment response, which has been particularly demonstrated with statin therapy for primary prevention of CAD, with significantly greater benefit in patients with high PRS for CAD (232,233,242,322). We investigated if genetic risk prediction with PRS may help to guide anticoagulation.

Using PRESTIGE-AF trial data of 128 participants, we evaluated the association of several PRS with incident IS, recurrent ICH risk, and secondary trial endpoints (i.e. major bleeding, net clinical benefit), and their potential to stratify DOAC efficacy and safety among ICH survivors with AF. We used previously validated PRS models available for AF (323) and ICH (213), and an integrative PRS for ischemic stroke (IS-iPRS) (10). We also included 'weighted-GRS' for WMH and ischemic stroke using only independent SNPs reaching genome-wide significance ($p_t<5x10^{-8}$). We tested the association of these PRS models with primary and secondary trial endpoints, including in interaction models with treatment arm (DOAC vs. Control), and evaluated three-year prediction performance in FGR models against established clinical risk scores (CHA2DS2-VASC and HAS-BLED). Additionally, we wanted to see if genetic risk modified the efficacy of DOACs, and assessed differences in absolute risk reduction due to DOAC in four groups: Control + Low PRS, Control + High PRS, DOAC + Low PRS, and DOAC + High PRS. We also investigated if index event bias impacted our results, leveraging data from the UK Biobank.

Though no PRS model was associated with primary or secondary endpoints in the overall trial, our analyses presented several interesting findings. Notably, an AF-PRS significantly interacted with treatment arm for overall stroke and net clinical benefit, and improved prediction of net clinical benefit by an estimated 10.1% over the CHA2DS2-VASc risk score. Genetic risk stratification further revealed that high AF-PRS patients derived the greatest benefit from DOACs to prevent stroke ($DOAC_{ARR}$=-20.0%, $p_{log\text{-}rank}$=0.02) and major adverse cardiovascular events ($DOAC_{ARR}$= -25.0%, $p_{log\text{-}rank}$=0.01). Among other PRS, the IS-iPRS predicted the net clinical benefit endpoint in the control group, and showed strong association with ischemic stroke and net clinical benefit after adjustment for index event bias. However, we are limited in interpreting these results since all ischemic events occurred in the control arm and most bleeds in the DOAC arm. Thus, our findings could reflect true effect modification or differential PRS associations between ischemic versus bleeding risk.

Overall, despite small sample limitations, we provide evidence that PRS, especially AF-PRS, may serve as a risk stratification tool to identify ICH survivors with AF who are most likely to benefit from anticoagulation with DOACs. This study has important clinical implications for the use of PRS in secondary prevention. PRS are not currently recommended in current guidelines for routine cardiovascular disease risk prediction (324), and there is a need for more definitive randomized trials to quantify the potential benefits of PRS in relevant clinical scenarios (325). High-stakes clinical scenarios like this—marked by severe outcomes and therapeutic ambiguity—may represent the most pragmatic starting points for PRS implementation in secondary stroke prevention.

These findings should be replicated in ongoing phase-3 trials for DOACs in ICH survivors with AF. There is also opportunity to leverage large-scale biobanks to train and optimize PRS models for prediction in clinical settings. Finally, this work demonstrates the immense value of collecting biological samples as a routine component of RCTs. We suggest potential utility of PRS in secondary stroke prevention especially when integrated with clinical risk prediction tools.

Article 2 (first author)

Tim D'Aoust*[5], Shufan Huo, Santiago Clocchiatti-Tuozzo, Peter U. Heuschmann[1,2], Joan Montaner[3,4], Jemma Hopewell, Morgane Lachaize[5,6], Guillaume Paré[7], Charles Wolfe[8, 9], Guido Falcone, Roland Veltkamp[10, 11, 12], Stéphanie Debette[5, 6]

Contribution of the PhD Candidate: all statistical analyses, interpretation of results, and writing of full manuscript.

The manuscript is currently being circulated to co-authors and will submitted soon, targeting *Circulation*. The circulated version is presented in the following pages and the supplementary material is available in **Appendix 7**.

Full Title:

Polygenic risk may inform anticoagulation treatment in intracerebral hemorrhage survivors with atrial fibrillation

Authors:

Tim D'Aoust*[1], Joan Montaner[2,3], Santiago Clocchiatti-Tuozzo[4,5], Jemma Hopewell[6], Guido Falcone[4,5], Morgane Lachaize[1,7], Guillaume Paré[8], Charles Wolfe[9,10], Peter U. Heuschmann[11,12], Roland Veltkamp[13, 14, 15], Stéphanie Debette[1, 7,16]

[1]Bordeaux Population Health Center, INSERM, UMR U1219, University of Bordeaux, 146 Rue Léo Saignat 11, 33076, Bordeaux, France

[2]Institute de Biomedicine of Seville, IBiS/Hospital Universitario Virgen del Rocío/CSIC/University of Seville & Department of Neurology, Hospital Universitario Virgen Macarena, Seville, Spain,

[3]Neurovascular Research Laboratory, Vall d'Hebron Institute of Research (VHIR), Hospital Vall d'Hebron, Barcelona, Spain,

[4]Department of Neurology, Yale School of Medicine, 333 Cedar St New Haven, CT 06510, USA.

[5]Center for Brain and Mind Health, Yale School of Medicine, 333 Cedar St New Haven, CT 06510,

[6]Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield Department of Population Health (P.R.G., A.V.E., F.M., A.O., C.F.C., R.S.W., J.C.H.), University of Oxford, United Kingdom.

USA.

[7]Department of Neurology, Institute for Neurodegenerative Diseases, CHU de Bordeaux, Pl. Amélie Raba Léon, 33000, Bordeaux, France,

[8]McMaster University & Hamilton Health Sciences. Population Health Research Institute (PHRI) and Thrombosis and Atherosclerosis Research Institute (TAaRI), Hamilton, Canada,

[9]Stroke Unit - Internal, Vascular and Emergency Medicine, University of Perugia, Santa Maria della Misericordia Hospital Perugia, Perugia, Italy,

[10]NIHR Applied Research Collaboration (ARC) South London, London, United Kingdom,

[11]Clinical Trial Center Würzburg, University Hospital Würzburg, Würzburg, Germany,

[12]Institute of Clinical Epidemiology and Biometry, University of Würzburg, Würzburg, Germany

[13]Department of Brain Sciences, Imperial College London, London, United Kingdom,
[14]Department of Neurology, Alfried-Krupp Krankenhaus, Essen, Germany, Essen, Germany,
[15]Department of Neurology, University Heidelberg, Heidelberg, Germany
[16]Institut du Cerveau (ICM), Paris Brain Institute, INSERM U1127, UMR CNRS 7225 Paris, Sorbonne Université, Assistance Publique des Hôpitaux de Paris, Paris, France

Short Title:

Polygenic risk scores to predict stroke risk in ICH survivors with AF

Corresponding author:

Stéphanie Debette, Bordeaux University, 146 rue Léo Saignat, Bordeaux Cedex 33076, France; Institut du Cerveau (ICM), Sorbonne Université, 47/83 boulevard de l'Hôpital, Paris 75013, France. Email: stephanie.debette@u-bordeaux.fr; stephanie.debette@icm-institute.org

Abstract

Background: Survivors of intracerebral hemorrhage (ICH) with atrial fibrillation (AF) are at high risk of both ischemic stroke and recurrent ICH. PRESTIGE-AF, a randomized controlled trial, showed DOACs substantially decrease the risk of ischemic stroke but also increase the risk of recurrent ICH. Polygenic risk scores (PRS) that identify individuals at highest genetic risk of complex traits could help inform clinical decision-making in this context.

Methods: PRS for AF, ischemic stroke (IS), intracerebral hemorrhage (ICH) and white matter hyperintensities (WMH), a highly prevalent risk factor for IS and ICH, were derived in 128 PRESTIGE-AF participants (65 DOAC, 63 control) who consented to a genetic substudy. Associations of PRS with incident IS, recurrent ICH, and any stroke were estimated using Cox regression, overall and by treatment arm, with analysis of interaction for any stroke. Prediction performance over clinical risk scores (CHA2DS2-VASc, HAS-BLED) was assessed, and survival analyses stratified across PRS tertiles were conducted. Secondary safety and net clinical benefit endpoints were investigated.

Results: Over a median follow-up of 1.3[0.7-2.1] years, 19 participants experienced primary stroke endpoints (11 IS; 8 ICH), and 25 major cardiovascular-related events in total. No PRS was significantly associated to stroke or secondary outcomes in the overall trial. We observed a significant interaction of AF-PRS with treatment arm for any stroke risk (p=0.02) and net clinical benefit (p=0.004), with opposite effect directions in DOAC patients and controls. AF-PRS showed significant improvement in AUC beyond CHA2DS2-VASc for net clinical benefit by +10.1% (p=0.03, AUC=0.77[0.63-0.90]). In stratified analyses, the absolute risk reduction (ARR) in any stroke risk attributable to DOAC was significant among those in the upper tertile of AF-PRS (plog-rank=0.02, DOACARR=-20%), but not among those in the lower tertile of AF-PRS. An integrative IS PRS (IS-iPRS) was nominally associated to net clinical benefit in the control group (HR=2.14[1.1-4.16],p=0.03).

Conclusions: PRS, particularly AF-PRS, could provide clinically meaningful stroke risk stratification for DOAC treatment in ICH survivors with AF. Future follow-up of these findings in additional ongoing DOAC trials in ICH survivors with AF is warranted.

Background

Direct oral anticoagulants (DOACs) are the preferred therapy for stroke prevention in atrial fibrillation (AF), reducing thromboembolic events with lower bleeding risk compared to vitamin K antagonists (e.g., warfarin)[1,2]. However, optimal stroke prevention strategies remain uncertain for AF patients with high bleeding risk, such as intracerebral hemorrhage (ICH) survivors. Approximately 20-25% of ICH survivors have AF and are often anticoagulated at the time of ICH: these patients lack clear guidance on resuming (or starting) anticoagulation, relying on individualized risk-benefit assessments. Intial meta-analysis of four early stage clinical trials in ICH survivors with AF revealed uncertain effects of oral anticoagulation on stroke risk and cardiovascular death[3]. More recently, the PRESTIGE-AF trial (NCT03996772) became the first completed, phase 3 randomized clinical trial (N=319) investigating DOAC efficacy and safety in ICH survivors (<12 months) with AF. The trial showed DOACs reduced ischemic stroke (IS) risk but increased major hemorrhagic complications, including recurrent ICH, without establishing non-inferiority for ICH risk[4]. A recent meta-analysis including PRESTIGE-AF and two of the earlier stage clinical trials, SoSTART and APACHE-AF, showed a significant increase in recurrent ICH and major bleeding without a significant reduction in ischemic stroke or systemic embolism[5].These mixed findings underscore the challenge of balancing ischemic and bleeding risks, highlighting the need for enhanced clinical risk prediction tools based on circulating biomarkers, neuroimaging, and/or genetic data to guide anticoagulation in ICH survivors with AF.

Stroke has a substantial genetic component, with heritability estimates ranging from 16.1-40% for IS, depending on sub-type, and up to 44% for ICH[6,7]. Recent genome-wide association studies (GWAS) have identified 89 independent stroke risk loci (mostly for IS and any stroke), from which polygenic risk scores (PRS) were derived for genetic risk prediction[8]. PRS, aggregating genetic burden across common variant effects, have shown considerable predictive capacity in cardiovascular disease[9], and have potential to enhance clinical trial efficiency through predictive and prognostic enrichment[10]. There is increasing evidence that PRS for stroke and associated traits may improve stroke risk prediction and stratification in clinical trials[8,11], and among those with existing clinical indications[12,13]. Recent studies have also suggested added value of PRS for predicting outcomes in IS survivors, for IS recurrence[14], blood pressure control[15], incident AF[16], and mortality[17], but evidence is lacking for stroke recurrence or post-stroke outcomes in ICH survivors. Importantly, PRS also has potential to inform treatment response, which has been particularly demonstrated with statin therapy for

primary prevention of coronary artery disease (CAD), with significantly greater benefit in patients with high PRS for CAD[18–21]. PRS utility in informing anticoagulation for secondary stroke prevention in patients at high risk of both ischemic and hemorrhagic stroke remains underexplored.

Using PRESTIGE-AF trial data, we conducted a genetic sub-study to assess the added value of PRS for informing the prescription of DOACs in patients with a history of both ICH and AF. We evaluated the relation of several PRS with incident IS, recurrent ICH risk, and secondary trial endpoints (i.e. major bleeding, net clinical benefit), and their potential to stratify DOAC efficacy and safety among ICH survivors with AF. We focused specifically on PRS for IS, AF, ICH, and white matter hyperintensities (WMH), the most common MRI-marker of cerebral small vessel disease known to portend an increased risk of both IS and ICH[22].

Methods

PRESTIGE-AF genetic sub-study description

The PRESTIGE-AF trial recruited 319 individuals, assigning 158 to the DOAC group and 161 to the no anticoagulant (control) group[4]. At each recruitment site, participants were asked if they wished to take part in the genetic sub-study at the time of main trial recruitment, and again at subsequent follow-up visits until the end of the trial (May 31, 2024). Samples were collected at their enrolment visit, or at follow-up visits if consent was provided later on, and then centralized at the Vall d'Hebron Research Institute (VHIR) in Barcelona, Spain. A total of 130 participants (40.8%) were included in the genetic sub-study.

Genotyping and Imputation

DNA was extracted from the buffy coat obtained from blood, and a back-up EDTA tube of whole blood was also drawn. Genome-wide genotyping was performed using the Affymetrix Precision Medicine Axiom Array at the Clinical Research Laboratory and Biobank – Genetic and Molecular Epidemiology Laboratory (CRLB-GMEL, Canada). Quality control and Haplotype Reference Consortium[23] imputation with the Helmholz Munich server[24] are described in the **supplement**. After quality control, genotype data was available for 128 participants (one participant was removed due to sex discrepancy, and one was removed due to excess heterozygosity).

Primary and Secondary Endpoints

We analyzed co-primary PRESTIGE-AF endpoints of incident IS and recurrent ICH, as well as any stroke (i.e., either IS or recurrent ICH). For secondary endpoints, we analyzed safety

endpoints of major bleeding according to the International Society on Thrombosis and Haemostasis (ISTH) classification[25], and a 'net clinical benefit' endpoint comprised of major adverse cardiovascular events (MACE): any stroke, systemic embolism, myocardial infarction, cardiovascular mortality, or major bleeding. Full outcome definitions are available in the **supplement.**

Polygenic Risk Scores

We downloaded previously validated PRS models available for atrial fibrillation[26] (available from: https://cvd.hugeamp.org/downloads.html#polygenic), and ICH[13] (PGS Catalog ID: PGS003457), and an integrative PRS for IS[8] (PGS Catalog ID: PGS002724). We also generated 'weighted-GRS' models using a standard clumping and thresholding approach[27] for WMH and IS including only independent SNPs reaching genome-wide significance (r^2<0.1; p_t < 5e-08).

To generate robust PRS in PRESTIGE-AF, we used only SNPs with an imputation score >0.8, a minor allele frequency (MAF) > 0.01, and a call rate of > 0.95. PRS were generated with the PRSice-2 software[28] using the 1000G-EUR reference panel[29] for linkage disequilibrium clumping and minor allele frequency imputation of missing genotypes. The general formula for calculating a polygenic risk score for each individual, *i*, is given as:

$$PRS_i \propto \sum_{j=1}^{m} \alpha_j X_{ij}$$

where *m* is the total number of SNPs in the PRS, α_j is the effect size for the *j-th* SNP in the PRS model (from GWAS), and X_{ij} is the number of risk alleles of the *i-th* individual for the *j-th* SNP (0, 1, or 2).

Statistical Analyses

Association Testing

We employed Cox proportional hazard models to test the association of PRS with primary stroke endpoints (incident IS, recurrent ICH, and any incident stroke) and secondary endpoints (major bleed and net clinical benefit). We used time since treatment initiation as the time scale for those in the DOAC (treatment) group, and time since baseline visit for those in the control group. For each outcome, participants were censored at the date of the event, death or the end of study follow-up. For Cox models for stroke sub-types (IS and ICH) participants was censored if they experienced a stroke of the other type. All models were adjusted for age at baseline visit, sex, index ICH localization (lobar or non-lobar), lag between index ICH event and start of the

trial, and first five principal components of population stratification (PCs). Models for any stroke or net clinical benefit were also adjusted for treatment arm (DOAC vs. Control), and not for IS, ICH, or major bleeds due to lack of model convergence (due to most or all cases of stroke subtypes or major bleeds occurring in one treatment group). Given that we are most interested in PRS to inform DOAC treatment, we also tested for interaction between PRS and treatment arm using any stroke or net clinical benefit outcomes, and report strata-specific hazard ratios (HR).

We conducted further sensitivity analyses to determine the robustness of our findings. First we further adjusted both interaction and main effect models by established clinical risk scores, either the CHA2DS2-VASC[30] (for ischemic stroke risk or other thromboembolic events) or the HAS-BLED[31] score (for major bleeding risk). We estimated PRS associations in Fine-Gray models to assess if there was a different effect on the cumulative incidence of primary and secondary outcomes and to comprehensively assess the influence of competing risks of death (or stroke of other sub-types). We also applied an inverse probability weighting approach (IPW) to investigate index event bias in PRESTIGE-AF patients[32]. Index event bias is a type of selection bias that occurs when a population in selected on the basis of previous occurrence of an event (e.g. the index ICH event). This selection process creates a spurious inverse association between independent risk factors of the index event, and tends to bias any estimation of the association of these risk factors with recurrence or disease progression towards the null[33,34]. IPW functions to attempts to address this bias by down-weighting overrepresented individuals (i.e., those most likely to be ICH survivors with AF) and up-weighting underrepresented individuals (i.e., those least likely to be ICH survivors with AF)[32,35]. IPW weights were estimated in the UK Biobank (237 ICH survivors with AF; N=407,462) to represent the probability of being an ICH survivor with AF in the general population. A summary of the models run for each outcome, and full description of the sensitivity analysis to account for potential index event bias is available in the **supplement.**

Prediction

We assessed three-year prediction performance of PRS for primary and secondary endpoints based on time-dependent AUC[36]. We used Fine-Gray regression models since this method is recommended for prediction over Cox proportional hazard models[37]. All models included the same covariates described for association testing, and interaction models (PRS by treatment arm) were used for any stroke and net clinical benefit endpoints. We compared prediction performance of PRS models beyond established clinical risk scores (CHA2DS2-VASC and

HAS-BLED), reporting % improvement in AUC and Wald test p-values[38]. Prediction performance was estimated using the *Score* function from the *riskRegression* R package[39].

Risk Reduction due to DOAC stratified by Genetic Risk

To stratify by genetic risk we categorized individuals based on PRS tertile contrasting those in the lower or upper tertile, and then constructed four groups based on treatment arm and genetic risk: Control + Low PRS, Control + High PRS, DOAC + Low PRS, and DOAC + High PRS. For this analysis, we selected robust PRS indicating association with primary and secondary outcomes. Before applying this stratification, PRS were adjusted for the baseline covariates in order to isolate genetic risk independent from these factors (age, sex, PCs, index ICH location, lag, and treatment arm). The residualized PRS were then used to determine genetic risk tertiles. We estimated the absolute risk reduction (ARR) attributable to DOAC within upper and lower PRS tertiles for both primary and secondary endpoints. We also constructed survival curves to compare probability of stroke over time in these groups, using a log-rank test to test differences in survival distributions between groups focusing on the DOAC effect within PRS tertiles.

Results

Description of the study population and endpoints

After genetic quality control (two participants removed), the PRESTIGE-AF genetic sub-study included 128 participants (40% of the original trial) with 63 and 65 assigned to the control (no anticoagulant) group and DOAC group respectively. Participants were balanced between groups for relevant variables such as mean age at enrolment, female sex, and location of index ICH (control: median age=79.4 years, 36.5% female, 33.3% lobar index ICH; DOAC: median age=77.1 years, 33.8% female, 40% lobar index ICH) (**Table 1**). In the control group, 11 (17.5%) patients experienced an IS, one (1.5%) an ICH recurrence. In the DOAC group, there were no IS cases, and seven patients had an ICH reccurence (10.8%). Overall, in the control group there were 15 (23.8%) major adverse cardiac events (MACE) contributing to the net clinical benefit endpoint: 12 events were ischemic (one systemic embolism), two bleeding-related, and one cardiovascular-related death due to myocardial infarction. In the DOAC group all 10 (15.4%) MACE were bleeding-related, including three major bleeds other than recurrent ICH, two of which were intracranial (non-ICH) hemorrhages (**Table 2**).

After quality control, 6,980,126 viable SNPs were available to derive PRS. Summary of the PRS models for IS, ICH, AF, and WMH can be found in **Supplementary Table 1**.

Association testing

Primary endpoints

No PRS was significantly associated with IS, ICH recurrence, or any stroke in models adjusted for age at baseline visit, sex, index ICH localization, lag between index ICH event and start of the trial, and first five genetic PCs (incl. treatment arm for any stroke) (**Table 3a**). We found significant interaction with treatment group for associations of AF-PRS with any stroke risk, with opposite effect directions in each treatment arm ($p_{interaction}$=0.02) (**Table 4a**).

Secondary endpoints

No PRS model was significantly associated with major bleeding or net clinical benefit in the overall trial (**Table 3b**). For the net clinical benefit endpoint, we did see significant interaction with treatment arm for the AF-PRS ($p_{interaction}$=0.004) and the ICH-PRS ($p_{interaction}$=0.01), with opposite effect directions in the DOAC and control arms (**Table 4b**). AF-PRS showed a significant protective association to this endpoint (HR_{per1SD}=0.29 [0.10-0.84], $p_{in\text{-}strata}$=0.02) in those anticoagulated with DOAC – only major bleeds were experienced in this group. Interaction with IS-iPRS was non-significant ($p_{interaction}$=0.08), but IS-iPRS did have significant association to the net clinical benefit endpoint – mostly ischemic events – in the control group (HR_{per1SD}=2.14[1.10-4.16], $p_{in\text{-}strata}$=0.03).

Sensitivity Analyses

All analyses were not sensitive to additionally adjusting for CHA2DS2-VASc or HAS-BLED (**Supplementary Tables 2-4**). Fine-Gray modeling to examine associations with cumulative incidence of stroke (subtypes) while accounting for competing risks (of death or alternative stroke subtypes) did not show substantial differences across most analyses, including analyses of interaction (**Supplementary Table 5-6**). Some subgroup analyses showing significance with Cox models were no longer statistically significant using Fine-Gray modeling, such as associations of the AF-PRS with cumulative incidence of ICH recurrence in the DOAC group (sHR_{per1SD}=0.30 [0.08-1.07], p=0.06), or IS-PRS association to cumulative incidence of net clinical benefit in the control group (sHR_{per1SD}=1.97[0.94-4.12], $p_{in\text{-}strata}$=0.07). These findings suggest marginal influence of competing risk (of death or alternative censored events).

To account for potential index event bias, IPW weights were derived for a set of stroke-related comorbidities and risk factors (age, sex, hypertension, diabetes, previous ischemic stroke, congestive heart failure and smoking status) in the selection model in UK Biobank (**Supplement**). After applying IPW, IS-iPRS showed a stronger and significant association with

ischemic stroke (HR_{per1SD} = 3.55 [1.40-9.00], p=0.008) and the net clinical benefit endpoint (HR_{per1SD} =2.24 [1.34-3.76], p=0.002) in the overall trial. There was no significant change in other associations between PRS and primary and secondary endpoints (**Supplementary Table 7**).

Prediction

Using Fine-Gray regression models, we aimed to assess the ability of PRS to predict primary and secondary endpoints above and beyond clinical risk scores at three-years. HAS-BLED was used for ICH recurrence and major bleeding and CHASDS2-VASc was used for ischemic stroke, any stroke, and net clinical benefit. WMH-GRS had the best performance for ischemic stroke (AUC=0.831 [0.625,1.00]), while AF-PRS improved prediction the most for recurrent ICH (AUC=0.75 [0.587,0.913]), any stroke (AUC=0.748 [0.602,0.895]), major bleeding (AUC=0.792 [0.612,0.966]), and net clinical benefit (AUC=0.765 [0.63,0.90]) (**Figure 1**). Compared to CHASDS2-VASc, AF-PRS significantly improved performance for the net clinical benefit endpoint, increasing AUC by +10.1% (95% CI [1.0-19.3%], p=0.03) **(Figure 1E)**. Improvement over clinical risk scores was non-significant for all other outcomes (**Supplementary Table 8a-e**).

Risk stratification by treatment and genetic risk groups

We stratified PRESTIGE-AF patients into four groups based on PRS tertile (lower and upper) and treatment arm: Control + Low PRS, Control + High PRS, DOAC + Low PRS, and DOAC + High PRS. For this anlaysis we selected IS-iPRS and AF-PRS, both of which showed significant associations or interactions for primary or secondary endpoints. We derived absolute risk reduction by DOAC across PRS tertiles and performed stratified survival analyses, testing the difference between outcome risk distributions across groups for each primary and secondary endpoint (**Supplementary Tables 9-10**).

Among patients with high AF-PRS (upper tertile), there were no ICH recurrences, and we saw a significant reduction of IS risk due to DOAC ($p_{log\text{-}rank}$=0.02, ARR= -20% [-42.3, 2.3%], **Figure 2A**). Among patients with high IS-iPRS, there was also a significant reduction in IS risk due to DOAC ($p_{log\text{-}rank}$=0.04, $DOAC_{ARR}$ =-22.7% [-45.0, -0.4]), however this is offset by a borderline significant trend towards a higher risk of ICH recurrence ($p_{log\text{-}rank}$=0.062, $DOAC_{ARR}$=+15% [-5.4, +35.4] (**Figure 2B)**. Accordingly, among patients with high AF-PRS there was a significant reduction in any stroke risk due to DOAC (log-rank p=0.024), which

was not observed among patients with high IS-iPRS ($p_{log\text{-}rank}$=0.661, $DOAC_{ARR}$=-7.7% [-36.0, +20.5).

Similar to primary stroke endpoints, DOAC significantly improved net clinical benefit among patients with high AF-PRS ($p_{log\text{-}rank}$=0.01, $DOAC_{ARR}$= -25.0%[-48.7,-1.3]) (**Figure 2A4**), with no major bleedings occurring in the DOAC + high PRS group. In contrast, we found no significant improvement of net clinical benefit due to DOAC in patients with high IS-iPRS ($p_{log\text{-}rank}$=0.21, $DOAC_{ARR}$=-21.4% [-51.6, +8.9].

Discussion

The PRESTIGE-AF trial showed that, in ICH survivors with AF, DOACs substantially decreased risk of ischemic stroke but consequently increased risk of recurrent ICH and major bleeding. In this genetic sub study, we found that PRS, particularly an AF-PRS derived from the largest AF-GWAS[26], may help identify ICH survivors with AF most likely to benefit from anticoagulation with DOACs. AF-PRS significantly interacted with treatment arm for overall stroke and net clinical benefit, showing opposite effect directions across treatment arms, and improved prediction of net clinical benefit by an estimated 10.1% over the CHA2DS2-VASc risk score. Genetic risk stratification further revealed that high AF-PRS patients derived the greatest benefit from DOACs to prevent stroke ($DOAC_{ARR}$=-20.0%, $p_{log\text{-}rank}$=0.02) and major adverse cardiovascular events ($DOAC_{ARR}$= -25.0%, $p_{log\text{-}rank}$=0.01). Among other PRS, the IS-iPRS, incorporating PRS for vascular risk factors and stroke sub-types across multiple ancestries, predicted net clinical benefit in the control group, and showed strong association with IS and net clinical benefit in the overall trial after adjustment for index event bias. These findings support the potential utility of PRS in refining risk stratification and guiding individualized anticoagulation strategies in ICH survivors with AF.

PRS has potential to inform treatment response and secondary prevention, as has been previously shown for β-blockers in heart failure patients[40,41], and alirocumab (PSCK9 inhibitor) in patients with CAD[20]. To our knowledge, no prior study has investigated if DOAC effectiveness may differ according to genetic risk. We observed significant interactions of AF-PRS and ICH-PRS with DOAC treatment. IS-iPRS was also associated with net clinical benefit in the control, but not the DOAC arm. Risk stratification analysis, in particular, further supported that DOACs may be especially effective in those with high AF-PRS. However, these findings could reflect true effect modification or important differential PRS associations between ischemic versus bleeding risk—all ischemic events occurred in the control arm and

most bleeds in the DOAC arm. We encourage further investigations of PRS, particularly AF-PRS, to modify treatment response to DOACs – this will be possible in future meta-analysis with other ongoing DOAC trials in ICH survivors with AF.

AF-PRS is a strong predictor of incident AF[26,42] and earlier AF onset[43], and has been associated to IS in AF patients beyond clinical risk factors[44]. In the UK Biobank, the AF-PRS used here was also associated with other vascular comorbidities (e.g., hypertension, peripheral vascular disease), and, most strongly, with other arrythmias and structural cardiac diseases (e.g. tricuspid valve disease)[26]. AF-PRS may also capture genetically-driven, more severe AF and cardiac dysfunction that confers higher IS risk. In PRESTIGE-AF, although directionality was consistent, we do not find a significant association of the AF-PRS with IS risk. However, AF-PRS significantly improved prediction of net clinical benefit beyond clinical risk scores, and notably, in the upper AF-PRS tertile, no bleeding events were observed, and DOACs markedly reduced absolute IS risk and improved net clinical benefit. Taken together, we suggest a high AF-PRS, conferring more 'genetically-driven' AF, may identify individuals with greater ischemic risk, and additionally, lesser vulnerability to bleeding. Individuals with low AF-PRS may alternatively reflect AF arising from greater systemic comorbidities that are independently associated with bleeding risk — such as hypertension, diabetes, or small vessel disease. Testing these hypotheses would require larger population-based and clinical samples to compare comorbidity burden across AF genetic risk groups.

Stroke PRS have shown ability to predict stroke in individuals with AF, including a 32-SNP GRS that identified increased stroke risk in anticoagulated patients with low CHA_2DS_2-VASc scores (≤ 2)[1]. In PRESTIGE-AF, a similar 33-SNP IS-GRS derived from the latest IS-GWAS[8] was not significantly associated with stroke risk overall in PRESTIGE-AF, and we could not investigate within patients with low CHA_2DS_2-VASc scores due to limited sample size (only 13 patients with CHA_2DS_2-VASc ≤ 2). The IS-iPRS (incorporating PRS for vascular risk factors and stroke sub-types across multiple ancestries) previously predicted IS risk in cross-ancestry biobanks and across five European clinical trials[8]. Here, the IS-iPRS was associated with net clinical benefit in the control but not in the DOAC arm, which may reflect that DOAC partially offsets genetic risk. In risk stratification, however, those with high IS-iPRS see a significant reduction in IS risk due to DOAC, but also higher risk of recurrent ICH. Compared to IS-iPRS, AF-PRS seems to provide better risk stratification to inform DOAC therapy in this highly selected patient population.

Overall, PRS were not directly associated with primary and secondary endpoints in the genetic substudy. Prior work has shown that AF-PRS and IS-iPRS predict ischemic stroke[8,12,26,44], and ICH-PRS predicts first ICH and bleeding under warfarin[13], but recurrence and progression studies often report attenuated or even inverse associations which are explained in part by selection bias[32–34,45]. Such attenuation has been observed in PRS studies of both recurrent coronary artery disease and stroke[8,12,46,47], although some large biobank studies demonstrate PRS can predict stroke recurrence[14,48]. Indeed our small, highly-selected sample had limited power and could have obscured PRS-outcome associations. Even so, we provide important evidence for clinical utility: PRS, at least marginally, improved overall prediction beyond clinical scores across endpoints, and AF-PRS significantly enhanced prediction of net clinical benefit.

We considered the possibility that index event bias impacted our findings by attenuating PRS-outcome associations[33]. To our knowledge only one other study has comprehensively considered index event bias for PRS-recurrent stroke associations[14]. Similarly, we applied an IPW-based approach to re-weight individuals, leveraging UK Biobank data to account for the probability of developing AF and surviving ICH. The adjusted results revealed a particularly strong association of IS-iPRS with IS (HR=3.55 [1.40-9.00], p=0.008) and net clinical benefit (HR=2.24 [1.34-3.76], p=0.002). This finding underscores that index event bias attenuated the effects of IS-iPRS, likely because its component PRS are strongly linked to both AF and ICH. Unexpectedly, the adjustment had little effect on ICH recurrence, suggesting our model was better at capturing selection factors related to AF than ICH risk. Still, our results highlight the importance of accounting for selection bias when studying recurrent events and suggest that IS-iPRS remains a meaningful predictor of ischemic outcomes in this population. Additional, larger studies in this context are needed to more comprehensively assess index event bias, especially for recurrent bleeding risk.

A key strength of this study is the use of genetic data from the first phase 3 RCT in ICH survivors with AF—a high-risk population where optimal treatment practices remains uncertain. Trial data offer the distinct advantages of randomization, standardized follow-up, and adjudicated outcomes, minimizing confounding and misclassification common in observational cohorts. In a highly relevant clinical context, we were able to assess the added value of PRS beyond conventional clinical scores for predicting stroke, bleeding, and net clinical benefit, and to explore whether genetic risk modifies DOAC effectiveness. We also

conducted robust sensitivity analyses, including IPW adjustment for index event bias– an important selection bias in this patient population.

The main limitation is sample size (128 participants, 40% of the PRESTIGE-AF trial), which restricted statistical power and precluded subgroup analyses (e.g., by CHA_2DS_2-VASc score or DOAC type). Due to the low number of outcomes, we were unable to formally test effect modification between a PRS and DOAC treatment for each stroke sub-type, or DOAC effect within genetic risk strata. Instead, we conducted non-parametric analysis to compare DOAC risk reduction within PRS tertile. We acknowledge that reported prediction metrics (AUCs) may be optimistic, as models were evaluated internally, highlighting the need for external validation. Additionally, although we attempted to account for index event bias using IPW, this approach assumes no unmeasured confounding and is not designed for prediction in highly selected populations. The PRESTIGE-AF trial overall has limited generalizability: women are under-represented (35% of participants), the trial excluded ICH survivors with severe disability, and all participants are of European ancestry. Additionally, while AF-PRS and IS-iPRS are based on large, multi-ancestry GWAS and validated across multiple settings, both the ICH-PRS and WMH-GRS are from European-ancestry GWAS with more limited power (especially for ICH), and need more refinement and validation. Thus, our findings have limited generalizability to non-European populations— a key barrier to clinical implementation of PRS[49].

This study has important clinical implications for the use of PRS in secondary prevention. PRS are not currently recommended in guidelines for routine CVD risk prediction[50], and there is a need for more definitive randomized trials to quantify the potential benefits of PRS in relevant clinical scenarios[51]. ICH survivors with AF face simultaneously high risks of ischemic and hemorrhagic stroke, and the benefit of DOAC treatment remains uncertain. Our study demonstrates that PRS can refine risk stratification in this context of clinical equipoise and could help identify optimal treatment strategies where uncertainty is greatest. High-stakes clinical scenarios like this—marked by severe outcomes and therapeutic ambiguity—may represent the most pragmatic starting points for PRS implementation in secondary stroke prevention.

We highlight a few avenues for future research. First, PRS would benefit from further refinement in this context, for example, by incorporating pharmacogenetic information capturing effects on drug metabolism that could modify drug efficacy and safety[52], or training PRS models in specific patient populations (e.g., ICH survivors with AF) identified within large-scale biobanks. Second, PRS should ultimately be implemented into clinical practice

through integration with existing clinical risk prediction tools, and, potentially, novel circulating biomarkers. For example, combining CAD-PRS with QRISK-2 has shown feasibility in UK primary care[53], and biomarkers such as NT-proBNP and GDF-15 have improved stroke prediction when added to CHA_2DS_2-VASc[54]. Studies of such integrated risk tools to inform anticoagulation therapy should be also investigated.

In conclusion, this genetic sub study of the PRESTIGE-AF trial underscores the value of collecting genetic data as a routine component of RCTs. We provide evidence that PRS may serve as a risk stratification tool to identify ICH survivors with AF who are most likely to benefit from anticoagulation with DOACs. Replication in ongoing phase-3 trials in ICH survivors with AF will be essential to validate these results and support their translation into clinical practice.

References

1. Joglar JA, Chung MK, Armbruster AL, et al. 2023 ACC/AHA/ACCP/HRS Guideline for the Diagnosis and Management of Atrial Fibrillation: A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. *Circulation*. 2024;149(1):e1-e156. doi:10.1161/CIR.0000000000001193

2. Van Gelder IC, Rienstra M, Bunting KV, et al. 2024 ESC Guidelines for the management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): Developed by the task force for the management of atrial fibrillation of the European Society of Cardiology (ESC), with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. Endorsed by the European Stroke Organisation (ESO). *Eur Heart J*. 2024;45(36):3314-3414. doi:10.1093/eurheartj/ehae176

3. Salman RAS, Stephen J, Tierney JF, et al. Effects of oral anticoagulation in people with atrial fibrillation after spontaneous intracranial haemorrhage (COCROACH): prospective, individual participant data meta-analysis of randomised trials. Accessed September 1, 2025. https://www.thelancet.com/journals/laneur/article/PIIS1474-4422(23)00315-0/fulltext

4. Veltkamp R, Korompoki E, Harvey KH, et al. Direct oral anticoagulants versus no anticoagulation for the prevention of stroke in survivors of intracerebral haemorrhage with atrial fibrillation (PRESTIGE-AF): a multicentre, open-label, randomised, phase 3 trial. *The Lancet*. 2025;405(10482):927-936. doi:10.1016/S0140-6736(25)00333-2

5. Kaisaier W, Guan C, Ye M, et al. Effect of oral anticoagulant therapy on adverse outcomes in patients with atrial fibrillation after intracranial hemorrhage. Accessed September 1, 2025. https://www.heartrhythmjournal.com/article/S1547-5271(25)02602-5/fulltext

6. Bevan S, Traylor M, Adib-Samii P, et al. Genetic Heritability of Ischemic Stroke and the Contribution of Previously Reported Candidate Gene and Genomewide Associations. *Stroke*. 2012;43(12):3161-3167. doi:10.1161/STROKEAHA.112.665760

7. Falcone GJ, Woo D. Genetics of Spontaneous Intracerebral Hemorrhage. *Stroke*. 2017;48(12):3420-3424. doi:10.1161/STROKEAHA.117.017072

8. Mishra A, Malik R, Hachiya T, et al. Stroke genetics informs drug discovery and risk prediction across ancestries. *Nature*. 2022;611(7934):115-123. doi:10.1038/s41586-022-05165-3

9. Abraham G, Rutten-Jacobs L, Inouye M. Risk Prediction Using Polygenic Risk Scores for Prevention of Stroke and Other Cardiovascular Diseases. *Stroke*. 2021;52(9):2983-2991. doi:10.1161/STROKEAHA.120.032619

10. Fahed AC, Philippakis AA, Khera AV. The potential of polygenic scores to improve cost and efficiency of clinical trials. *Nat Commun*. 2022;13(1):2922. doi:10.1038/s41467-022-30675-z

11. O'Sullivan JW, Ashley EA, Elliott PM. Polygenic risk scores for the prediction of cardiometabolic disease. *Eur Heart J*. 2023;44(2):89-99. doi:10.1093/eurheartj/ehac648

12. Marston NA, Patel PN, Kamanu FK, et al. Clinical Application of a Novel Genetic Risk Score for Ischemic Stroke in Patients With Cardiometabolic Disease. *Circulation*. 2021;143(5):470-478. doi:10.1161/CIRCULATIONAHA.120.051927

13. Mayerhofer E, Parodi L, Prapiadou S, et al. Genetic Risk Score Improves Risk Stratification for Anticoagulation-Related Intracerebral Hemorrhage. *Stroke*. 2023;54(3):791-799. doi:10.1161/STROKEAHA.122.041764

14. Kojima N, Koido M, He Y, et al. Recurrent Stroke Prediction by Applying a Stroke Polygenic Risk Score in the Japanese Population. *Stroke*. Published online June 2025. doi:10.1161/STROKEAHA.124.047786

15. Acosta JN, Both CP, Demarais ZS, et al. Polygenic Susceptibility to Hypertension and Blood Pressure Control in Stroke Survivors. *Neurology*. 2023;100(15):e1587-e1597. doi:10.1212/WNL.0000000000206763

16. Tack RW, Tan BYQ, Senff JR, et al. Predicting Atrial Fibrillation After Stroke by Combining Polygenic Risk Scores and Clinical Features. *Stroke*. 0(0). doi:10.1161/STROKEAHA.124.050123

17. Li J, Chaudhary D, Griessenauer CJ, Carey DJ, Zand R, Abedi V. Predicting mortality among ischemic stroke patients using pathways-derived polygenic risk scores. *Sci Rep*. 2022;12(1):12358. doi:10.1038/s41598-022-16510-x

18. Natarajan P, Young R, Stitziel NO, et al. Polygenic Risk Score Identifies Subgroup With Higher Burden of Atherosclerosis and Greater Relative Benefit From Statin Therapy in the Primary Prevention Setting. *Circulation*. 2017;135(22):2091-2101. doi:10.1161/CIRCULATIONAHA.116.024436

19. Marston NA, Kamanu FK, Nordio F, et al. Predicting Benefit From Evolocumab Therapy in Patients With Atherosclerotic Disease Using a Genetic Risk Score. *Circulation*. 2020;141(8):616-623. doi:10.1161/CIRCULATIONAHA.119.043805

20. Damask A, Steg PG, Schwartz GG, et al. Patients With High Genome-Wide Polygenic Risk Scores for Coronary Artery Disease May Receive Greater Clinical Benefit From Alirocumab Treatment in the ODYSSEY OUTCOMES Trial. *Circulation*. 2020;141(8):624-636. doi:10.1161/CIRCULATIONAHA.119.044434

21. Klarin D, Natarajan P. Clinical utility of polygenic risk scores for coronary artery disease. *Nat Rev Cardiol*. 2022;19(5):291-301. doi:10.1038/s41569-021-00638-w

22. Debette S, Schilling S, Duperron MG, Larsson SC, Markus HS. Clinical Significance of Magnetic Resonance Imaging Markers of Vascular Brain Injury: A Systematic Review and Meta-analysis. *JAMA Neurol*. 2019;76(1):81-94. doi:10.1001/jamaneurol.2018.3122

23. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet*. 2016;48(10):1279-1283. doi:10.1038/ng.3643

24. Das S, Forer L, Schönherr S, et al. Next-generation genotype imputation service and methods. *Nat Genet*. 2016;48(10):1284-1287. doi:10.1038/ng.3656

25. Schulman S, Kearon C. Definition of major bleeding in clinical investigations of antihemostatic medicinal products in non-surgical patients. *J Thromb Haemost*. 2005;3(4):692-694. doi:10.1111/j.1538-7836.2005.01204.x

26. Roselli C, Surakka I, Olesen MS, et al. Meta-analysis of genome-wide associations and polygenic risk prediction for atrial fibrillation in more than 180,000 cases. *Nat Genet*. 2025;57(3):539-547. doi:10.1038/s41588-024-02072-3

27. Choi SW, Mak TSH, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. *Nat Protoc*. 2020;15(9):2759-2772. doi:10.1038/s41596-020-0353-1

28. Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data. *GigaScience*. 2019;8(7):giz082. doi:10.1093/gigascience/giz082

29. Auton A, Abecasis GR, Altshuler DM, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74. doi:10.1038/nature15393

30. Lip GYH, Nieuwlaat R, Pisters R, Lane DA, Crijns HJGM. Refining Clinical Risk Stratification for Predicting Stroke and Thromboembolism in Atrial Fibrillation Using a Novel Risk Factor-Based Approach. Accessed September 1, 2025. https://journal.chestnet.org/article/S0012-3692(10)60067-0/abstract

31. Pisters R, Lane DA, Nieuwlaat R, de Vos CB, Crijns HJGM, Lip GYH. A novel user-friendly score (HAS-BLED) to assess 1-year risk of major bleeding in patients with atrial fibrillation: the Euro Heart Survey. *Chest*. 2010;138(5):1093-1100. doi:10.1378/chest.10-0134

32. Mitchell RE, Hartley AE, Walker VM, et al. Strategies to investigate and mitigate collider bias in genetic and Mendelian randomisation studies of disease progression. *PLOS Genet*. 2023;19(2):e1010596. doi:10.1371/journal.pgen.1010596

33. Smits LJM, Kuijk SMJ van, Leffers P, Peeters LL, Prins MH, Sep SJS. Index event bias—a numerical example. Accessed September 2, 2025. https://www.jclinepi.com/article/S0895-4356(12)00283-1/fulltext

34. Dahabreh IJ, Kent DM. Index Event Bias as an Explanation for the Paradoxes of Recurrence Risk Research. Accessed September 2, 2025. https://jamanetwork.com/journals/jama/fullarticle/645707

35. Cole SR, Hernán MA. Constructing Inverse Probability Weights for Marginal Structural Models. *Am J Epidemiol*. 2008;168(6):656-664. doi:10.1093/aje/kwn164

36. Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat Med*. 2013;32(30):5381-5397. doi:10.1002/sim.5958

37. Austin PC, Fine JP. Practical recommendations for reporting Fine-Gray model analyses for competing risk data. *Stat Med*. 2017;36(27):4391-4400. doi:10.1002/sim.7501

38. Blanche P, Proust-Lima C, Loubère L, Berr C, Dartigues J, Jacqmin-Gadda H. Quantifying and Comparing Dynamic Predictive Accuracy of Joint Models for Longitudinal Marker and Time-to-Event in Presence of Censoring and Competing Risks. *Biometrics*. 2015;71(1):102-113. doi:10.1111/biom.12232

39. Gerds TA, Ohlendorff JS, Ozenne B. *RiskRegression: Risk Regression Models and Prediction Scores for Survival Analysis with Competing Risks*.; 2023. https://CRAN.R-project.org/package=riskRegression

40. Lanfear DE, Luzum JA, She R, et al. Polygenic Score for β-Blocker Survival Benefit in European Ancestry Patients With Reduced Ejection Fraction Heart Failure. *Circ Heart Fail*. 2020;13(12):e007012. doi:10.1161/CIRCHEARTFAILURE.119.007012

41. Lanfear DE, Luzum JA, She R, et al. Validation of a Polygenic Score for Beta-Blocker Survival Benefit in Patients With Heart Failure Using the United Kingdom Biobank. *Circ Genomic Precis Med*. 2023;16(2):e003835. doi:10.1161/CIRCGEN.121.003835

42. Marston NA, Garfinkel AC, Kamanu FK, et al. A polygenic risk score predicts atrial fibrillation in cardiovascular disease. *Eur Heart J*. 2023;44(3):221-231. doi:10.1093/eurheartj/ehac460

43. Mars N, Koskela JT, Ripatti P, et al. Polygenic and clinical risk scores and their impact on age at onset and prediction of cardiometabolic diseases and common cancers. *Nat Med*. 2020;26(4):549-557. doi:10.1038/s41591-020-0800-0

44. O'Sullivan JW, Shcherbina A, Justesen JM, et al. Combining Clinical and Polygenic Risk Improves Stroke Prediction Among Individuals With Atrial Fibrillation. *Circ Genomic Precis Med*. 2021;14(3):e003168. doi:10.1161/CIRCGEN.120.003168

45. Dudbridge F, Allen RJ, Sheehan NA, et al. Adjustment for index event bias in genome-wide association studies of subsequent events. *Nat Commun*. 2019;10(1):1561. doi:10.1038/s41467-019-09381-w

46. Howe LJ, Dudbridge F, Schmidt AF, et al. Polygenic risk scores for coronary artery disease and subsequent event risk amongst established cases. *Hum Mol Genet*. 2020;29(8):1388-1395. doi:10.1093/hmg/ddaa052

47. Thompson PL, Hui J, Beilby J, et al. Common genetic variants do not predict recurrent events in coronary heart disease patients. *BMC Cardiovasc Disord*. 2022;22(1):1-9. doi:10.1186/s12872-022-02520-0

48. Han Y, Shi H, Yu C, et al. Genetic Risk and Prognosis of the First Incident Stroke Survivors. *Neurology*. 2025;105(2):e213832. doi:10.1212/WNL.0000000000213832

49. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Current clinical use of polygenic scores will risk exacerbating health disparities. *Nat Genet*. 2019;51(4):584-591. doi:10.1038/s41588-019-0379-x

50. Visseren FLJ, Mach F, Smulders YM, et al. 2021 ESC Guidelines on cardiovascular disease prevention in clinical practice: Developed by the Task Force for cardiovascular disease prevention in clinical practice with representatives of the European Society of Cardiology and 12 medical societies With the special contribution of the European

Association of Preventive Cardiology (EAPC). *Eur Heart J*. 2021;42(34):3227-3337. doi:10.1093/eurheartj/ehab484

51. Schunkert H, Di Angelantonio E, Inouye M, et al. Clinical utility and implementation of polygenic risk scores for predicting cardiovascular disease: A clinical consensus statement of the ESC Council on Cardiovascular Genomics, the ESC Cardiovascular Risk Collaboration, and the European Association of Preventive Cardiology. *Eur Heart J*. 2025;46(15):1372-1383. doi:10.1093/eurheartj/ehae649

52. Zhai S, Zhang H, Mehrotra DV, Shen J. Pharmacogenomics polygenic risk score for drug response prediction using PRS-PGx methods. *Nat Commun*. 2022;13(1):5278. doi:10.1038/s41467-022-32407-9

53. Fuat A, Adlen E, Monane M, et al. A polygenic risk score added to a QRISK®2 cardiovascular disease risk calculator demonstrated robust clinical acceptance and clinical utility in the primary care setting. *Eur J Prev Cardiol*. 2024;31(6):716-722. doi:10.1093/eurjpc/zwae004

54. Short SAP, Wilkinson K, Hald E, et al. Improving stroke risk prediction in atrial fibrillation with circulating biomarkers: the CHA2DS2-VASc–Biomarkers model. *J Thromb Haemost*. Published online August 6, 2025. doi:10.1016/j.jtha.2025.06.007

Tables and Figures

Table 1. Baseline characteristics of PRESTIGE-AF Genetic sub-study

Table 1. Baseline Characteristics of PRESTIGE-AF Genetic Sub-Study

	DOAC (N=65)	CONTROL (N=63)
Age at enrolment	77.1 (70.6-82.7)	79.4 (73.2-83.5)
Sex		
Male	43 (66.2%)	40 (63.5%)
Female	22 (33.8%)	23 (36.5%)
Medical History		
Hypertension	63 (96.9%)	60 (95.2%)
Diabetes	24 (36.9%)	16 (25.4%)
Myocardial infarction	6 (9.2%)	9 (14.3%)
Ischemic stroke	10 (15.4%)	15 (23.8%)
Transient ischemic attack	3 (4.6%)	3 (4.8%)
ICH	4 (6.2%)	5 (7.9%)
Other major bleeding	9 (13.8%)	8 (12.7%)
Modified Rankin Scale score	3 (1-3)	3 (1-3)
CHA2DS2-VASc score	4.7 (1.7)	4.8 (1.5)
HAS-BLED score	2.9 (0.9)	2.9 (0.9)
Blood pressure at enrolment, mmHg		
Systolic	127 (13.0)	127 (14.6)
Diastolic	75.7 (10.7)	76.5 (8.3)
Time from index ICH to treatment initiation, days	54 (35-100)	49 (29-94)
Location of ICH		
Lobar	26 (40.0%)	21 (33.3%)
Non-lobar	39 (60.0%)	42 (66.7%)
Antithrombotic medication before index ICH		
Anticoagulant agent	57 (87.7%)	55 (87.3%)
Antiplatelet agent	8 (12.3%)	5 (7.9%)
Both	2 (1.6%)	2 (1.6%)

Table 2. Outcomes in PRESTIGE-AF Genetic sub-study

	CONTROL (N=63)	**DOAC (N=65)**
Major Adverse Cardiac Events	15 (23.8%)	10 (15.4%)
Ischemic Stroke	11 (17.5%)	0
ICH	1 (1.6%)	7 (10.8%)
Intracranial hemorrhage (non-ICH)	0	2 (3.1%)
Other Major Bleed	1 (1.6%)	1 (1.5%)
Systemic Embolism	1 (1.6%)	0
CV-Related Death	1 (1.6%)	0
Non-CV Death	1 (1.6%)	1 (1.5%)

Table 3a. Associations between PRS with stroke endpoints in Cox models.

	Any Stroke		Ischemic Stroke		Recurrent ICH	
PRS	**HR[95% CI]**	**P**	**HR[95% CI]**	**P**	**HR[95% CI]**	**P**
AF-PRS	0.94 [0.57-1.54]	0.79	1.30 [0.67-2.53]	0.43	0.44 [0.17-1.13]	0.09
IS-iPRS	1.38 [0.80-2.37]	0.24	1.31 [0.63-2.71]	0.47	1.50 [0.66-3.40]	0.34
ICH-PRS	0.92 [0.52-1.64]	0.78	1.24 [0.57-2.69]	0.59	0.55 [0.21-1.42]	0.21
IS-GRS	1.06 [0.63-1.77]	0.83	1.42 [0.72-2.82]	0.31	0.81 [0.36-1.79]	0.59
WMH-GRS	1.21 [0.75-1.96]	0.43	1.81 [0.91-3.60]	0.09	0.83 [0.40-1.72]	0.61

Models were adjusted for age at baseline visit, sex, index ICH localization (lobar or non-lobar), lag between index ICH and start of trial, and first five principal components for population stratification (and treatment arm for any stroke).

Table 3b. Associations between PRS with major bleeding and net clinical benefit endpoints in Cox models.

	Major Bleeds		Net Clinical Benefit	
PRS	**HR[95% CI]**	**P**	**HR[95% CI]**	**P**
AF-PRS	0.47 [0.22-1.00]	0.05	0.81 [0.52-1.28]	0.37
IS-iPRS	1.37 [0.71-2.66]	0.35	1.43 [0.90-2.26]	0.13
ICH-PRS	0.92 [0.52-1.64]	0.14	0.91 [0.55-1.50]	0.70
IS-GRS	1.27 [0.67-2.39]	0.46	1.14 [0.73-1.77]	0.57
WMH-GRS	0.87 [0.49-1.56]	0.64	1.19 [0.80-1.79]	0.40

Models were adjusted for age at baseline visit, sex, index ICH localization (lobar or non-lobar), lag between index ICH and start of trial, and first five principal components for population stratification (and treatment arm for net clinical endpoint).

Table 4a. Interaction between PRS and treatment arm for risk of any stroke in Cox models.

PRS	P interaction	Arm	HR [95% CI]	P strata
AF-PRS	**0.02**	CONTROL	1.28 [0.75-2.19]	0.36
		DOAC	**0.29 [0.09-0.91]**	**0.03**
IS-iPRS	0.73	CONTROL	1.50 [0.72-3.10]	0.28
		DOAC	1.24 [0.56-2.78]	0.60
ICH-PRS	0.06	CONTROL	1.38 [0.65-2.93]	0.41
		DOAC	0.46 [0.18-1.21]	0.12
IS-GRS	0.14	CONTROL	1.42 [0.76-2.66]	0.27
		DOAC	0.66 [0.30-1.47]	0.31
WMH-GRS	0.55	CONTROL	1.36 [0.74-2.49]	0.32
		DOAC	1.01 [0.47-2.16]	0.98

Models were adjusted for age at baseline visit, sex, index ICH localization (lobar or non-lobar), lag between index ICH and start of trial, and first five principal components for population stratification. Strata-specific hazard ratios are reported by reparameterizing the cox model.

Table 4b. Interaction between PRS and treatment arm for the net clinical benefit endpoint in Cox models.

PRS	P interaction	Arm	HR [95% CI]	P strata
AF-PRS	**0.004**	CONTROL	1.17 [0.72-1.89]	0.52
		DOAC	**0.24 [0.09-0.64]**	**0.005**
IS-iPRS	0.08	**CONTROL**	**2.14 [1.10-4.16]**	**0.03**
		DOAC	0.93 [0.47-1.86]	0.85
ICH-PRS	**0.01**	CONTROL	1.51 [0.77-2.96]	0.24
		DOAC	**0.43 [0.19-0.98]**	**0.04**
IS-GRS	0.87	CONTROL	1.17 [0.66-2.07]	0.59
		DOAC	1.09 [0.55-2.16]	0.81
WMH-GRS	0.39	CONTROL	1.38 [0.81-2.35]	0.24
		DOAC	0.96 [0.51-1.81]	0.91

Models were adjusted for age at baseline visit, sex, index ICH localization (lobar or non-lobar), lag between index ICH and start of trial, and first five principal components for population stratification. Strata-specific hazard ratios are reported by reparameterizing the cox model.

Figure 1. Risk prediction of each trial endpoint at 3-years. Reference model includes baseline covariates. Each other model includes baseline covariates in addition to either a clinical risk score or PRS.

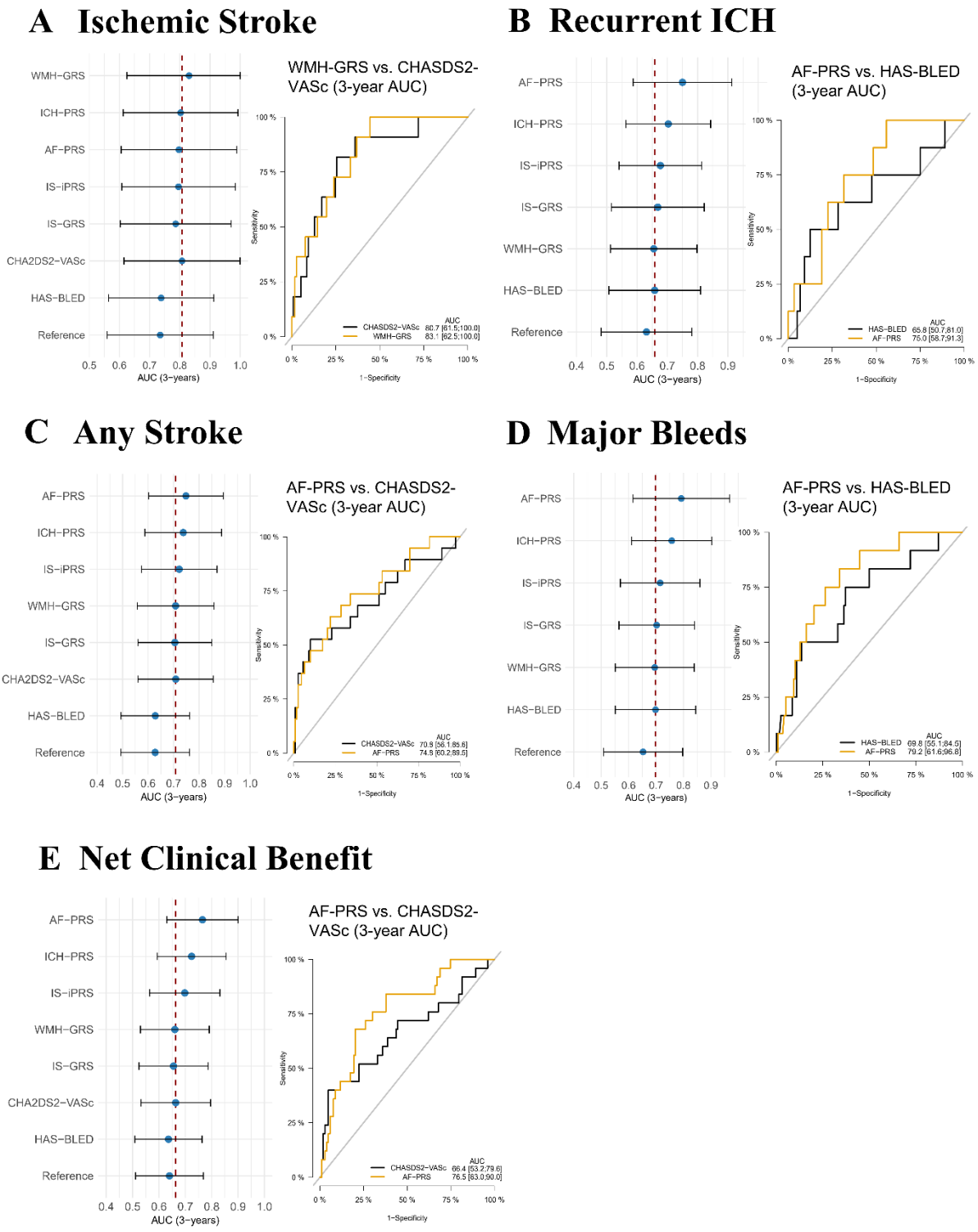
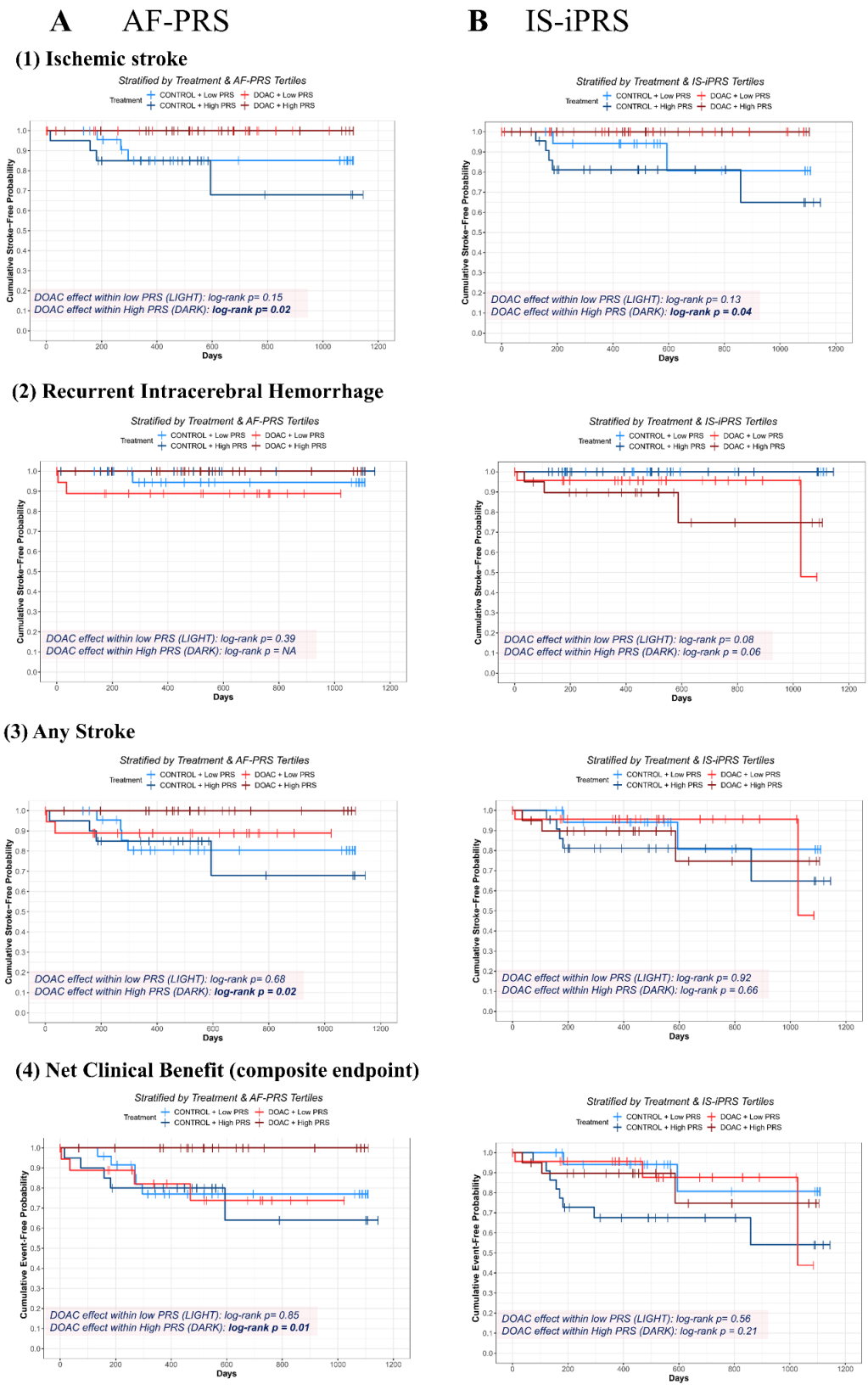


Figure 2. Risk stratification by PRS tertile and treatment group in non-paramtric Kaplan-meier analysis. A log-rank test to test differences in survival distributions between groups.



Chapter 6 - Pathway-based PRS approach to cSVD

In this third part of my thesis project, I apply both a standard genome-wide and pathway-based approach to WMH to optimize risk prediction, and notably, investigate cSVD etiology. The conception of this project was inspired from a few sources. First, a prior postdoctoral fellow with Pr Stéphanie Debette, Quentin Le Grand, proposed a project investigating WMH in i-Share – a French cohort of university students – as it is unknown whether WMH in young adults represents similar lesions in older adults related to cSVD or other conditions such as multiple sclerosis. This investigation was made feasible by the recent improvements in WMH quantification algorithms in individuals with sparse white matter lesions by colleageus as part of the RHU-SHIVA project (326). Originally, we planned to investigate PRS derived from GWAS for blood pressure traits, WMH, and MS to see if they were associated with WMH in younger adults. In addition to this project, at the annual RHU-SHIVA meeting there was a lot of discussion around the need to integrate multi-omics findings at the pathway level. Recently, a novel and easily implementable pathway-based PRS approach, *PRSet* was developed (314), and had been already applied to MS (311). Since cSVD is heterogenous and mechansims are largely unknown, a pathway-based PRS approach could be especially useful for delineating potential mechanisms and allow us to investigate PRS for these pathways in other independent cohorts (e.g., young adults). Thus, the project evolved to encompass a comprehensive PRS approach to optimize PRS for WMH and uncover underlying biological pathways with an added emphasis on a lifespan approach. This work was co-supervised by Quentin Le Grand and Pr Stéphanie Debette.

6.1 Pathway-specific polygenic risk scores reveal underlying mechanisms of cerebral small vessel disease

WMH, the cSVD MRI-marker with the largest GWAS meta-analysis, has 27 known risk loci representing genes involved in extracellular matrix (ECM) and cell membrane structure, myelination, and membrane transport (16). WMH-GRS have already been linked to dMRI markers of white matter microstructure in young adults, suggesting early genetic influences on disease pathogenesis (13,71). Further, genetically-predicted WMH showed evidence for a causal association with stroke and dementia outcomes in two-sample mendelian randomization

(MR) studies (13). Yet, MR does not make use of individual data and, unlike PRS, it cannot be used to investigate associations in independent cohorts.

WMH-PRS has not yet been optimized for prediction for WMH itself or for clinical cSVD complications, stroke and dementia. Furthermore, a pathway-based PRS approach may be particularly useful for cSVD where underlying etiology remains unclear. The aim of this work was to first optimize a WMH global-PRS to quantify overall genetic liability, and second, to apply an agnostic WMH pathway-specific PRS (ps-PRS) approach to identify canonical pathways implicated in cSVD etiology.

We first leveraged data from 15,320 UKBB participants with brain imaging (UKBB-MRI) to optimize a WMH global-PRS, as well as identify a set of WMH ps-PRS representing canonical pathways defined in existing pathway databases. Using 10-fold cross-validation, we evaluated 16 candidate WMH global-PRS models derived using a classic C+T approach and PRS-CS. The best WMH global-PRS model (PRS-CS, Φ=auto) was selected based on the minimization of the root-mean-squared-error.

To identify a set of pathways, we split the UKBB-MRI into two equally sized datasets, and with *PRSet* performed pathway enrichment analysis on 3,794 canonical pathways in the MSigDB v7.1 weighted based on effect sizes from the WMH-GWAS. We identified 127 WMH ps-PRS enriched in both subsets of UKBB-MRI. We performed Walktrap clustering based on the correlation of these WMH ps-PRS and identified ten clusters representing coherent biological processes, for which we also derived 'cluster-based' WMH ps-PRS. Then, we performed pathway enrichment of these WMH ps-PRS in three follow-up cohorts representing the adult life span and clinical spectrum: 3C-Dijon, i-Share, and Memento. We found 61 WMH ps-PRS to be enriched in at least one follow-up cohort. This set of 61 WMH ps-PRS and the WMH global-PRS were carried forward for secondary analyses.

Secondary analyses included investigating:

1. Association with stroke and dementia in the UKBB, the full 3C study, and Memento (dementia only). We included sub-types ICH and IS, and AD and vascular or mixed dementia.
2. Effect modification by hypertension for the association of WMH-PRS with WMH volume in UK Biobank.

First, the WMH global-PRS, was associated with WMH volume across all follow-up cohorts. We also originally demonstrated that WMH global-PRS is predictive of stroke (both IS and

ICH sub-types) and VMD, but not AD. WMH global-PRS, on the whole, appears to capture genetic risk for vascular contributions to dementia independent of AD.

We focused on WMH ps-PRS showing lifespan evidence and clinical relevance, and identified a set of biological pathways that may underlie cSVD hallmarks, including endothelial dysfunction, impaired BBB function, altered fluid drainage, and vascular inflammation (59,61). Pathways enriched in younger adults may point to early processes in cSVD development, while those found exclusively in older adults and memory-clinic patients could reflect more advanced pathology predisposing to stroke or dementia. Protein localization and fatty acid metabolism pathways emerged with the most robust evidence, particularly related to omega-3 and omega-6 fatty acids, with WMH ps-PRS enriched across the lifespan and associated with stroke and dementia. These findings align with observational and metabolomic studies linking hyperlipidemia and α-linolenic (omega-3) and linoleic (omega-6) fatty acid metabolites with WMH volume (327–329). Pathways related to organelle biosynthesis and maintenance, and specifically to ciliogenesis, also showed strong evidence across the lifespan, particularly those involved in basal body anchoring during primary cilia assembly. Notably, primary cilia in vascular endothelial cells mediate responses to shear stress from blood flow via nitric oxide and calcium channels (330,331). Dysfunction of these organelles can lead to aberrant fluid-sensing and contribute to hypertension and atherosclerosis (331–333), and also disrupt vascular barriers (e.g., the BBB) (334). We found consistent evidence across our analyses for cluster WMH ps-PRS for *signal transduction*, and multiple pathways involving signaling by *receptor tyrosine kinases* such as VEGF, platelet-derived growth factor (PDGF), and ephrin (EPH). We found several other WMH ps-PRS nominally associated with multiple stroke and dementia sub-types (e.g., *NPP1* and *Rho GTPase* pathways) and others that are particularly notable in the memory-clinic (e.g., sphingolipids and peroxisomes).

We first demonstrate that WMH global-PRS has potential utility for cSVD-related risk prediction. Through WMH-psPRS, we identify pathways related to lipid metabolism, ciliogenesis, and signal transduction, newly described in this study, that may underlie cSVD across the lifespan. We also suggest potential pathways to target therapeutically, including lipid subclasses (omega-3/6 fatty acids, sphingolipids, peroxisomes), NPP1 in the ECM, and tyrosine kinase signaling, particularly EPHA2. In addition to standard PRS approaches, we advocate for the potential of ps-PRS to guide top down biomarker discovery, complementing other approaches in multi-omics and experimental biology, especially for heterogenous diseases such as cSVD. Ps-PRS can also uniquely identify individuals at highest risk for pathway-specific

susceptibility, potentially informing personalized prevention and clinical trial enrichment. This work has potential to advance personalized cSVD prevention and offers a framework for the comprehensive use of PRS for not only prediction, but to better understand disease etiology as well.

Article 3 (first author)

Tim D'Aoust (MSc, Université de Bordeaux) ; Carole Dufouil (PhD, Université de Bordeaux) ; Ami Tsuchida (PhD, Université de Bordeaux) ; Quentin Le Grand* (PhD, Université de Bordeaux) ; Stéphanie Debette* (PhD, Université de Bordeaux)

Contribution of the PhD Candidate: all statistical analyses, interpretation of results, and writing of full manuscript.

The manuscript is currently being circulated to co-authors and will submitted soon, targeting *Brain*. The circulated version is presented in the following pages and the supplementary material is available in **Appendix 8**.

Pathway-specific polygenic risk scores reveal underlying mechanisms of cerebral small vessel disease

Tim D'Aoust (MSc, Université de Bordeaux) ; Carole Dufouil (PhD, Université de Bordeaux) ; Ami Tsuchida (PhD, Université de Bordeaux) ; Quentin Le Grand* (PhD, Université de Bordeaux) ; Stéphanie Debette* (PhD, Université de Bordeaux)

* These authors jointly supervised this work

Abstract

White matter hyperintensities (WMH), a key MRI-marker of cerebral small vessel disease (cSVD), are common in older adults and increase risk of stroke and dementia. The most recent WMH-GWAS identified 27 loci enriched for genes involved in extracellular matrix, myelination, and membrane transport. Genetically-predicted WMH have been linked to changes in white matter microstructure already in young adults and have shown causal effects on stroke (ischemic and hemorrhagic) and dementia. However, biological pathways underlying WMH, and their contribution to clinical outcomes, remain unclear. We applied global and pathway-specific PRS (ps-PRS) to optimize prediction and identify biological pathways contributing to WMH across the lifespan.

We leveraged the largest European-ancestry WMH-GWAS (N=46,944) and data from 15,320 UK Biobank participants with MRI (UKBB-MRI, mean age=67±7.7) to optimize a global WMH-PRS and construct 3,794 ps-PRS based on canonical pathways (MSigDB v2023.2). The best WMH global-PRS ($P_{UKBB\text{-}MRI}=1.78x10^{-193}$; $\Delta R^2=+4.1\%$) predicted WMH in independent cohorts: young adults (i-Share, N=1,578, age=22±2.3; P=0.0024; $\Delta R^2=+0.5\%$), older community-dwellers (3C-Dijon, N=1,443, age=73±4.1; $P=1.38x10^{-9}$; $\Delta R^2=+2.2\%$), and memory-clinic patients (Memento, N=1,831, age=71±8.5; $P=1.11x10^{-19}$; $\Delta R^2=+3.3\%$). In the larger UKBB sample (N up to 355,180), WMH global-PRS was associated with both stroke types (ischemic stroke, $HR_{IS}=1.061$[1.037-1.086] $P=1.6x10^{-8}$; intracerebral hemorrhage, $HR_{ICH}=1.083$[1.026-1.143], P=0.004) and vascular/mixed dementia ($HR_{VMD}=1.168$[1.103-1.237], $P=1.1x10^{-7}$), but not Alzheimer's disease.

Permutation-based pathway enrichment, performed in the UKBB-MRI sample, identified 127 ps-PRS consistently enriched for WMH and clustering into ten biological domains. Of these, 61 ps-

PRS (six clusters and 55 individual pathways) were enriched in at least one follow-up cohort: 14 in older community-persons, 17 in young adults, and 37 in memory-clinic patients.

Secondary analyses highlighted four ps-PRS involved in lipid metabolism, ciliogenesis, or signal transduction enriched in both young and older adults and associated with stroke and dementia. Some ps-PRS, notably sphingolipid metabolism, showed particular clinical relevance in memory-clinic patients. Seven ps-PRS, mostly lipid-related, showed evidence of modulation by hypertension.

In summary, we demonstrate the predictive utility of a global WMH-PRS for cSVD and introduce a multi-cohort ps-PRS framework that reveals candidate biological pathways relevant to WMH and cSVD across the lifespan. These findings provide novel insights into cSVD mechanisms, with potential implications for drug development and personalized prevention.

Introduction

Cerebral small vessel disease (cSVD), characterized by pathological changes in the structure and function of small brain vessels, is a known cause of lacunar ischemic stroke (IS) and intracerebral hemorrhage (ICH) and the main pathological substrate underlying the vascular contribution to cognitive decline and dementia[1]. Most often, cSVD manifests as 'covert' vascular lesions, detectable on brain magnetic resonance imaging (MRI) in the absence of clinical symptoms. Covert cSVD is extremely prevalent in older adults and is associated with increased risk of stroke, dementia, depression, and mortality, representing a major target to prevent age-related brain diseases in the general population[2]. Cardiovascular risk factors, particularly hypertension, are associated with cSVD, but only account for approximately 2% of the variance in cSVD lesion burden[3], suggesting a substantial role of genetic and other factors. While optimal management of cardiovascular risk factors can slow down the progression of cSVD[2], there are currently no mechanism-based treatments targeting the disease process.

The most studied MRI-marker of cSVD (MRI-cSVD) is white matter hyperintensities (WMH) of presumed vascular origin in the periventricular or deep white matter[4]. Histologically, they are heterogenous, representing varying degrees of ischemic demyelination, axonal loss, and gliosis[5]. Recent developments now enable precise quantification of WMH using automated algorithms in large-scale population-based cohorts, even in those with very small WMH volumes (e.g. young adults)[6]. WMH volume is associated with cSVD-related complications, primarily stroke, dementia[4], depression[7–9], and cognitive performance[10,11].

Genome-wide association studies (GWAS) of MRI-cSVD markers have begun to unravel molecular mechanisms underlying cSVD[12–15] identifying 27 genetic loci associated with WMH volume[13], with an overrepresentation of genes involved in extracellular matrix (ECM) and cell membrane structure, myelination, and membrane transport[12] . Polygenic risk scores (PRS) derived from GWAS aggregate the effects of numerous independent risk variants across the genome to capture global genetic risk for complex traits, and have demonstrated ability to identify individuals at high genetic risk of multifactorial diseases[16]. Interestingly, PRS focusing on genome-wide significant risk variants of WMH (also referred to as weighted genetic risk score, GRS), have been linked to additional cSVD MRI-markers of white matter microstructure, including diffusion-based metrics (e.g. diffusion tensor imaging (DTI), including peak width of skeletonized mean diffusivity

(PSMD), and neurite orientation dispersion and density imaging (NODDI) metrics) in young adults, suggesting early genetic influences on disease pathogenesis[13,17]. While such genome-wide PRS have been particularly useful in linking genetic determinants of WMH in older adults to other (early) imaging endophenotypes of cSVD susceptibility, their ability to predict cSVD complications has not been thoroughly explored. Furthermore, PRS representing 'global' genomic risk (global-PRS) are not specifically designed to provide information on disease etiology, and aggregating genetic risk according to functional genomic sub-structures may allow to better characterize the biological pathways involved in cSVD across the lifespan[18].

Unlike global-PRS, which provide a single, global genetic risk estimate, pathway-specific PRS (ps-PRS) partition genetic liability into biologically meaningful pathways. Ps-PRS can be defined with a priori knowledge of pathological processes (e.g. ps-PRS for Aβ clearance or cholesterol metabolism[19]) or agnostically[20,21], using data-driven approaches with pathway enrichment tools like MAGMA[22] or canonical databases to define viable pathways[23–26]. This hypothesis-free approach has uncovered enriched pathways involved in neurological conditions such as Parkinson's disease[21], multiple sclerosis[27], or AD-related changes in brain structure[20]. The advantage of ps-PRS over traditional pathway enrichment methods lies in its ability to obtain an estimate of an individual's susceptibility to specific disease-related pathways. This enables investigation of associations between enriched pathways and disease complications or biomarkers across independent cohorts representing different life stages and clinical contexts. For cSVD, where existing knowledge of well-defined pathways is limited, ps-PRS approaches could uncover novel biological pathways and provide important context for the role of putative biomarkers and therapeutic targets across the age and disease spectrum.

Leveraging biobank-scale data and a large GWAS meta-analysis of WMH in older adults, we first optimized a WMH global-PRS to quantify overall genetic liability. Second, we applied an agnostic WMH ps-PRS approach to identify canonical pathways implicated in cSVD etiology. In order to capture cSVD progression across the lifespan, we then attempted to replicate both WMH global-PRS and ps-PRS models across diverse cohorts: young adults, community-dwelling older adults, and memory-clinic patients. In secondary analyses we tested the association of the most robust ps-PRS with incident stroke and dementia outcomes in the general population and memory-clinic settings, and probed modulation of WMH (ps-)PRS by hypertension.

Materials and methods

An overview of our study design is provided in **Figure 1**.

Study populations

We leveraged multiple independent cohorts representing community and clinic-based populations across the lifespan. Across all cohorts we removed participants with pre-existing neurological conditions, missing information on WMH volume or genetic data, and individuals of non-European ancestry (see **Supplementary Methods** for details on exclusion criteria).

The UK Biobank is a prospective, population-based cohort of over 500,000 individuals with genetic data recruited between the ages of 40 and 69 between 2006 and 2010, and followed up longitudinally, with a large imaging sub-study (up to 100,000 participants). In our analysis of WMH volume, we focused on a subset of 15,320 European ancestry UK Biobank participants with brain MRI and available white matter hyperintensity volume, recorded after January 26th, 2020, to avoid overlap with participants included in the WMH GWAS meta-analysis used to derive the PRS. After exclusions (Supplementary Methods) this resulted in 15,320 eligible participants (mean-age=67.1±7.2 years, 53.2% women). Data from these were used to optimize genome-wide WMH PRS and identify a set of validated ps-PRS representing canonical biological pathways enriched for WMH.

The Three Cities (3C) study is a population-based cohort comprising community-dwelling persons aged $\geq$65 years from three French cities, Dijon, Montpellier, and Bordeaux, recruited in 1999-2001 and followed every 2-3 years over 12-17 years[28]. In Dijon, Participants less than age 80 enrolled between June 1999 and September 2020 were invited to undergo brain MRI, yielding 1,924 MRI scans[29]. After exclusions (**Supplementary Methods**) our sample comprised 1,443 eligible participants (mean age=72.7±4.1 years, 60.2% women).

The internet-based Students HeAlth Research Enterprise (i-Share) is a French prospective population-based study of university students aged 18-35 years[30], in which a sub-set of participants were invited for brain MRI and genome-wide genotyping through the MRi-Share and bio-Share ancillary studies[31,32]. Of 1,784 participants with brain MRI and genetic data[31], after exclusions (**Supplementary Methods**) our sample comprised 1,578 eligible participants (mean age=22.1±2.3 years, 71.7% women).

The Memento study is a memory clinic-based cohort[33], for which 2,246 participants with MCI or subjective cognitive complaints were recruited from 26 French memory clinics in 2011-2014, with brain MRI and a follow-up every 6-12 months over 5 years[33]. After exclusions (**Supplementary Methods**) we were left with 1,831 eligible participants (mean-age=71.1±8.5 years, 62.2% women).

MRI Acquisition and WMH Derivation

Brain MRI scans were obtained with a 1.5 Tesla (3C: Siemens Magnetom) or 3 Tesla (UK Biobank: Siemens Skyra; i-Share: Siemens Prisma) scanner and images were processed using a standardized protocol blinded to clinical or demographic features, as detailed elsewhere[29,32,34,35]. For Memento, 1.5 or 3 Tesla scans were collected at memory clinics and harmonized by the Center for Automated Treatment of Images (CATI; cati-neuroimaging.com)[36]. A fully automated software to quantify WMH volume from T1-, T2-weighted, and FLAIR (except 3C-Dijon) images was implemented in each study[6,37–39]. Total WMH volume was rank-based inverse-normal transformed to normalize and standardize the trait, and allow for better comparison across studies. In the i-Share cohort, due to the large number of equally-ranked values, we used indirect inverse-normal transformation of WMH volume, i.e., applying inverse-normal transformation on WMH-residuals, adjusted for sex, age at MRI, and the first four principal components of population stratification (PCs) (similar variables were adjusted for in the regression models for the other cohorts). More detailed description of MRI acquisitions and WMH derivation in each study is available in the **Supplementary Methods.**

Genotyping, quality control, and imputation

Genome-wide genotyping was performed using the Affymetrix UK BiLEVE Axiom Array and UK Biobank Axiom Array in UK Biobank, Affymetrix Precision Medicine Axiom Array in i-Share, and the Illumina Human610Iquad BeadChips in 3C-Dijon and in Memento. Quality control and imputation procedures are described elsewhere[31,40] and in the **Supplementary Methods**.

WMH Polygenic Risk Score Derivation

We used the largest WMH GWAS summary statistics comprising 46,944 older adults of European-ancestry to derive both WMH global-PRS and ps-PRS. These summary statistics were re-generated from the original study excluding 3C-Dijon to avoid sample overlap[13]. In addition, we removed all single nucleotide polymorphisms (SNPs) that were duplicated, multi-allelic, rare variants (minor allele frequency < 1%), or not available or mismatched in the 1000 Genomes European reference

panel[41]. For analyses in each target cohort respectively, SNPs were also removed if they were mismatched or had poor imputation quality (imputation score < 0.8).

WMH global-PRS

To first optimize WMH global-PRS, we derived 16 candidate WMH PRS models in 15,320 UK Biobank participants using different parameterizations for two established PRS methods: 'clumping and thresholding'[42], and PRS-CS[43] (**Figure 1A**, PRS methods description in **Supplementary Methods**). PRS models were calculated using the GenoPred pipeline which provides a streamlined, standardized framework to perform quality control and derive PRS in target datasets[44]. The 1000 Genomes European sample was used as an external linkage disequilibrium (LD) reference for both methods[41].

We implemented 10-fold cross validation using the *caret* R package[45] in order to select the WMH PRS that best predicts WMH volume in the UK Biobank. All models for testing association with inverse-normal transformed WMH volume (field 25781, instance 2) were adjusted for age at MRI (field 21003, instance 2), age at MRI^2, sex (field 31), total intracranial volume (field 26521), 10 PCs (field 22009), and genotyping chip (field 22000). The best WMH PRS was selected based on minimization of the root-mean-squared-error, and used as the WMH global-PRS in all subsequent analyses. We then tested the association of the WMH global-PRS with WMH volume in the 3C-Dijon, i-Share, and Memento studies, followed by meta-analysis across cohorts using the inverse-variance weighted (IVW) method.

Pathway-specific WMH PRS Discovery

We aimed to identify a set of ps-PRS that were enriched for WMH in the UK Biobank using the PRSet function implemented in PRSice-2 (v2.3.5)[42]. In an agnostic approach, we defined ps-PRS using gene sets from 3,794 curated biological pathways in the Canonical Pathway subset of Molecular Signatures Database (MsigDB version 7.4)[46], and weights from the latest WMH GWAS[13]. To validate enriched pathways, we split 15,320 UK Biobank participants randomly into two sub-samples, UKBB-I and UKBB-II (**Figure 1B**). We used a permutation-based, empirical p-value (p_{emp}) of < 0.05 in UKBB-I first to filter candidate pathways and then validated these pathways in UKBB-II to select ps-PRS enriched for WMH. More details on the PRSet pathway enrichment are presented in the **Supplementary Methods**.

We then performed clustering using the *igraph* R package[47,48] to identify relevant clusters of correlated pathways. Specifically, Walktrap clustering[49] was performed on a graph network built from an adjacency matrix of correlations ($r>0.4$) between WMH ps-PRS. For clusters of more than three ps-PRS, representing a sufficient number of non-identical pathways, we defined 'cluster ps-PRS' using expanded gene sets to include all genes within pathways in the same cluster. We inputted the gene sets of these cluster ps-PRS into g:Profiler[50] to perform functional profiling based on molecular functions, biological processes, and cellular components from Gene Ontology[51,52]. We assigned relevant names to each cluster based on the most significant Gene Ontology annotations from the full list of genes in the cluster, and confirmed enrichment of cluster ps-PRS in UKBB-I and UKBB-II with PRSet.

Next, we aimed to replicate pathway enrichment of individual and cluster WMH ps-PRS enriched in the UK Biobank ($p_{emp}<0.05$) in the 3C-Dijon, i-Share, and Memento studies. Linear regression was also used in each study to obtain direct effect estimates for association of WMH ps-PRS with WMH volume, which were then meta-analyzed across studies using IVW. All pathway enrichment and linear regression analyses were adjusted for the same covariates as WMH global-PRS derivation, and cohort-specific variables as necessary (study centre in 3C-Dijon, memory-clinic in Memento). PRS were mean-zero standardized, and all associations were reported per 1SD of PRS.

Secondary Analyses

For secondary analyses, we selected the WMH global-PRS, six replicating cluster ps-PRS and 55 individual ps-PRS enriched in at least one follow-up cohort.

Association of WMH PRS with cSVD complications, stroke and dementia

To analyze the association of enriched WMH (ps-)PRS with incident stroke and dementia, we used all samples with genetic data available in the UK Biobank (without brain MRI to avoid sample overlap), 3C, and Memento (for dementia only) studies (**Supplementary Methods**). We investigated associations of WMH-(ps-)PRS with any stroke, intracerebral hemorrhage (ICH) and ischemic stroke (IS), and with all-cause dementia, Alzheimer's disease (AD) and vascular or mixed dementia (VMD). For analyses of any incident stroke cases, we removed all prevalent stroke cases, and for analyses of stroke sub-types (IS or ICH), we additionally removed those with 'undefined' stroke to avoid misclassification bias. Similar exclusions were applied to incident dementia

analyses. For incident stroke, study samples comprised 349,283 UK Biobank participants (9,212 incident stroke cases; 7,281 IS; 1,344 ICH), and 5,745 3C participants (232 incident stroke cases; 184 IS; 42 ICH). For incident dementia, we analyzed 355,180 UK Biobank participants (7,924 incident all-cause dementia (ACD) cases; 2,728 AD; 1,187 vascular or mixed dementia), 5,883 3C participants (662 incident all-cause dementia cases; 459 AD; 131 vascular or mixed dementia), and 2,064 Memento patients (287 incident all-cause dementia cases; 197 AD; 37 vascular or mixed dementia).

We tested the associations of WMH (ps-)PRS with incident stroke and dementia using Cox proportional hazards models. We used age as a time scale, from age at baseline visit to end-of-follow-up, censoring at death, end of follow-up or event of other sub-type (for sub-type models). All models were adjusted for sex, genetic PCs and other study-specific covariates (e.g., study centre). For association analyses with dementia, models were additionally adjusted for *APOE-ε2* and *APOE-ε4* dosages. We adapted models in each cohort to meet Cox proportional hazard model assumptions. Clinical outcome ascertainment and study-specific exclusions and modelling decisions are fully described in the **Supplementary Methods**.

Effect of hypertension on the association of WMH ps-PRS with WMH volume

In the UK Biobank MRI sample (UKBB-I and UKBB-II), we also sought to identify WMH (ps-)PRS whose association with WMH volume is modulated by hypertension, the strongest risk factor for cSVD.

For this analysis, we include all participants from the initial UK Biobank MRI sample with blood pressure information (blood pressure levels and medication intake) at the time of MRI (N= 12,520, 81.7% of main analytical sample). We aimed to see if the associations of WMH global-PRS or ps-PRS are modulated by hypertension status. Hypertension was defined by systolic blood pressure > 140 mmHg or diastolic blood pressure > 90 mmHg or anti-hypertensive medication intake (**Supplementary Methods**). We divided the sample into sub-groups based on hypertension status and estimated the association of WMH PRS (pathway-specific and global) with WMH volume among hypertensives and non-hypertensives using linear regression models adjusted for the aforementioned covariates. Cochran's Q test for heterogeneity was used to determine effect modification for WMH (ps-)PRS by hypertension status.

Results

Baseline characteristics of the UK Biobank, 3C-Dijon, Memento, and i-Share studies are shown in **Table 1.**

WMH global-PRS

In 15,320 UK Biobank participants with brain-MRI, the PRS-CS-auto model was best able to predict WMH in 10-fold cross-validation, and was strongly associated with WMH volume (Beta=0.204 95%CI [0.190-0.217]; $p=1.78x10^{-193}$, variance explained (R^2) over base model=+4.1%). The predictive ability of WMH global-PRS was sensitive to both PRS method and parameterization with R^2 ranging from 1.8-4.1% over the base model with covariates only (Base model R^2=26.9%, **Supplementary Table 1**). We used the PRS-CS model (ϕ=auto) to represent WMH global-PRS in all subsequent analyses.

We obtained the associations of WMH global-PRS with WMH volume in each follow-up cohort using linear regression models adjusted for the same covariates. The WMH global-PRS was significantly associated to WMH volume in 3C-Dijon (Beta=0.153 [0.104-0.202], $p=1.38x10^{-9}$, R^2= +2.2%), Memento (Beta=0.197 [0.155-0.239], $p=1.11x10^{-19}$, R^2= +3.3%), and, to a lesser extent but significantly, in the much younger i-Share cohort (Beta=0.079 [0.028-0.130], p=0.0024, R^2=+0.5%). IVW meta-analysis across all four cohorts yielded a significant association between WMH global-PRS and WMH volume (Beta=0.161 [0.106-0.216], $p=1.74x10^{-8}$).

WMH ps-PRS

In UKBB-I (n=7,660), we identified 369 canonical pathways (out of 3,794) enriched for WMH (P_{emp}<0.05). In UKBB-II, 127 WMH of these 369 ps-PRS were also enriched and carried forward in pathway clustering and follow-up analyses in independent cohorts. All 127 WMH ps-PRS were enriched in the direction of increased WMH volume. Using Walktrap clustering on the 127 WMH ps-PRS we identified ten clusters of more than three ps-PRS encompassing 63/127 enriched pathways (49.6%) **(Figure 2, Supplementary Table 2)**. Full enrichment results for all 127 individual and 10 cluster ps-PRS in each cohort are included in **Supplementary Tables 3-6**.

No ps-PRS cluster was enriched in all three follow-up cohorts, but enrichment for six out of ten ps-PRS clusters was observed in at least one independent follow-up cohort (**Figure 3**). The *protein localization and fatty acid metabolism* cluster (comprising 8 correlated pathways and 247 genes) was enriched in both memory-clinic patients and young adults. The *organelle biogenesis* cluster (8

pathways, 415 genes) was enriched in older community-persons, the *signal transduction* (4 pathways, 623 genes) and *carbohydrate metabolism (disorders)* cluster (4 pathways, 623 genes) were enriched in young adults, and the *immune response to viral infections* (4 pathways, 1019 genes) and *metabolism and biosynthesis of nucleotide sugars and glycolysis* (7 pathways, 111 genes) clusters were enriched in memory-clinic patients. Although the *extracellular matrix, sphingolipid metabolism,* and *inflammatory signalling* did not replicate at the cluster level, several individual ps-PRS comprising the cluster did (**Figure 3**).

At the individual ps-PRS level, 55/127 ps-PRS (of which 26 within and 29 outside clusters) were enriched in at least one follow-up cohort: 14 in older community-persons, 14 in young adults, and 38 in memory-clinic patients **(Figure 3)**. Three pathways were enriched in both older community-persons and memory-clinic patients relating to *microtubule depolymerization* and *elastic fiber* pathways. Two pathways were enriched in both young and older community-dwelling adults: *anchoring of the basal body to the plasma membrane* and *alpha linolenic omega-3 and linoleic omega-6 metabolism.* One pathway, *ephrin type-A receptor 2 forward signaling,* was enriched in memory-clinic patients and young adults.

Finally, we estimated the associations of the 61 replicating WMH ps-PRS (six clusters and 55 individual ps-PRS) with WMH volume in each cohort using linear regression models adjusted for the same covariates, and performed IVW meta-analysis across all cohorts (including UK Biobank). Using Bonferroni multiple testing correction ($p<0.00105$; 49 independent tests determined in UKBB-MRI), 35 WMH ps-PRS were significantly associated to WMH volume in meta-analysis, showing consistent association across all cohorts (**Supplementary Tables 7&8**).

Secondary Analyses

In addition to WMH global-PRS, we included the six clusters and 55 individual ps-PRS that were enriched in at least one follow-up cohort in secondary analyses for association with stroke or dementia in cohorts of older adults, and for effect modification by hypertension in UK Biobank. Results of secondary analyses for prioritized pathways are summarized in **Figure 4**.

Association of WMH PRS with cSVD clinical complications, namely stroke and dementia

Descriptions of 3C, UK Biobank, and Memento samples used in these analyses are presented in **Supplementary Table 9**.

The WMH global-PRS was significantly associated with incident any stroke (HR_{AS}=1.06 [1.038-1.082], P=2.6x10^{-8}) and its sub-types (HR_{IS}=1.061 [1.037-1.086], P=1.6x10^{-8}; HR_{ICH} =1.083 [1.026-1.143], P=0.004) in 349,283 UK Biobank participants. For dementia, the global WMH-PRS was associated with incident ACD (HR_{ACD}=1.051 [1.03-1.08], P=1x10^{-5}) and more strongly with VMD (HR_{VMD}=1.168 [1.103-1.237], P=1.1x10^{-7}) in 355,180 UK Biobank participants, but with incident AD (HR_{AD}=1.018 [0.98-1.057], P=0.35). While the WMH global-PRS did not reach significance in association with any clinical outcomes in the much smaller 3C or Memento studies (N=2,064 to 5,883), associations were directionally consistent.

Several clusters showed nominal associations with incident stroke or dementia, the strongest associations (p<0.01) being observed between *organelle biosynthesis* and VMD or between *metabolism of nucleotide sugars and glycolysis* with ACD.

Of the 55 replicated WMH ps-PRS, 27 were nominally associated with incident stroke (P<0.05), and 32 with incident dementia outcomes in either population-based (UK Biobank or 3C) or memory-clinic (Memento) cohorts (**Supplementary Figure 1**). Detailed results of all associations between WMH ps-PRS with clinical outcomes can be found in **Supplementary Table 10-14**.

Some interesting patterns are worth highlighting. Several ps-PRS appear to be associated (p<0.01) with incident ischemic stroke specifically, namely *R signalling by PDGF*, *BC NPP1 pathway*, *KG small lung cancer*, and two Rho GTPase related pathways (*R Rho GTPase cycle* and *R signalling by Rho GTPases MIRO GTPases and RhoBTB3)*. The latter two were also associated (p<0.01) with vascular and mixed dementia. Interestingly, the *BC NPP1 pathway* and *KG small lung cancer* pathways had replicated with WMH in young adults, suggesting lifespan effects. One ps-PRS related to *galactose metabolism pathway*, which had replicated in a memory-clinic setting (Memento) showed significant association with all-cause dementia (p<0.01) and AD (p<0.001). Another ps-PRS, related to *chylomicron clearance*, showed strong association with ICH (p<0.001).

Effect modification by hypertension in the association of WMH-(ps-)PRS with WMH volume

We investigated whether the association of WMH-PRS was modified by hypertension status in UK Biobank participants. We tested for heterogeneity of associations between WMH-(ps-)PRS and WMH volume according to hypertension status (9,106 hypertensive; 3,414 non-hypertensive). The WMH global-PRS did not show notable differences in association between hypertension sub-

groups, and most WMH ps-PRS showed consistent effect sizes as well (**Supplementary Figure 2**, **Supplementary Table 15**). Seven WMH ps-PRS and clusters displayed nominally significant heterogeneity according to hypertension status ($p<0.05$), each showing stronger effects in the hypertensive group: the *protein localization and import and fatty acid metabolism* cluster ($p_{het}=0.028$) and two pathways within this cluster, *omega-3 and omega-6 fatty acid synthesis* ($p_{het}=0.002$) and *protein localization* ($p_{het}=0.033$); the *carbohydrate metabolism (disorders)* cluster ($p_{het}=0.024$), *chylomicron clearance* ($p_{het}=0.022$), and two pathways within the signal transduction cluster: *FCG receptor dependent phagocytosis* ($p_{het}=0.032$), and *parasite infection* ($p_{het}=0.032$).

Discussion

In this large multi-cohort study, leveraging global and ps-PRS for WMH derived from biobank-level data, we provide novel insights into the genetic underpinnings and biological pathways underlying WMH, the most-studied cSVD MRI-marker. We first confirmed that a global WMH PRS is associated with WMH volume in UK Biobank, with strong replication in three independent cohorts across the adult lifespan and clinical spectrum. Second, we identified 127 WMH ps-PRS enriched for WMH volume that clustered into ten coherent biological processes. These were validated across independent cohorts of young adults, community-dwelling older persons, and memory-clinic patients. Enriched WMH pathway clusters included notably those related to extracellular matrix and immune response, in line with previous findings in genomic and proteomic studies, as well as new pathways such as lipid metabolism, ciliogenesis, and signal transduction. Several clusters and individual WMH ps-PRS showed significant and distinct association patterns with incident stroke (ischemic or hemorrhagic) and dementia, particularly vascular and mixed subtypes. This work advances our understanding of WMH pathophysiology, identifying potential pathway-level mechanisms relevant for personalized cSVD prevention and treatment across the lifespan.

There is emerging evidence that processes predisposing to cSVD take root early in life[3,13,15,17,53]. Here we highlight several biological pathways that are enriched for WMH in both young and older adults with, notably, additional evidence for associations of several of them with stroke or dementia. These pathways converge on known cSVD hallmarks, including endothelial dysfunction, impaired blood–brain barrier (BBB) function, altered fluid drainage, and vascular inflammation[54,55].

Protein localization and fatty acid metabolism pathways emerged with the most robust evidence as enriched across the lifespan, both at the cluster and individual ps-PRS level, particularly related to omega-3 and omega-6 fatty acids. The alpha linolenic omega-3 and linoleic omega-6 acid metabolism was also associated with higher risk of incident ICH and lower risk of incident IS and VMD in the UK Biobank, at nominal significance, suggesting possible utility in stratifying risk of cSVD complications if validated independently. These findings align with observational and metabolomic studies linking α-linolenic (omega-3) and linoleic (omega-6) fatty acid metabolites with WMH volume[56–58]. Omega-3 and omega-6 fatty acids are polyunsaturated fatty acids (PUFAs)

that come from the diet, and are involved in cell membrane structure, inflammation, and vascular and brain health.[59] Interestingly, hypertension appeared to modulate effects of these ps-PRS (at nominal significance), consistent with recent Mendelian randomization-based mediation analyses suggesting that hypertension is the strongest mediator the causal effect of hyperlipidemia on WMH[60]. Dysregulated lipid metabolism is central to low-grade systemic inflammation in metabolic syndromes (e.g., obesity) and may underlie dysfunctional communication between the brain and the periphery found in these conditions[61]. Our findings thus reinforce lipid, and predominantly, fatty acid metabolism, as a possible central mechanism in cSVD across the lifespan.

Pathways related to organelle biosynthesis and maintenance, and specifically to ciliogenesis, also showed strong evidence across the lifespan, particularly those involved in *basal body anchoring* during primary cilia assembly. Several cilia-related and organelle biogenesis pathways were also enriched for WMH in older adults and associated with stroke and dementia. Overall, the organelle biosynthesis and maintenance (cilium) cluster was associated with VMD and, to a lesser extent, ischemic stroke. Primary cilia function as sensory organelles, and in vascular endothelial cells they mediate responses to shear stress from blood flow via nitric oxide and calcium channels[62,63]. In neuronal and glial cells, they mediate signaling cascades for vascular-endothelial growth factor (VEGF), nitric oxide, and amyloid-β peptides. Dysfunction of these organelles can lead to aberrant fluid-sensing and contribute to hypertension and atherosclerosis[63–65], and also disrupt vascular barriers (e.g., the BBB)[66]. Interestingly, cilia proteins have been proposed as novel biomarkers of damaged endothelium[62]. Moreover, *TRIM47*, the putative causal gene at the most significant WMH risk locus on chromosome 17q25[13,67], was recently shown to exert its effects by modulating the NRF2 pathway[68], NRF2 being a transcription factor reported to control cilia formation and function[69]. Our findings suggest that defects in primary cilia may play a central role in cSVD.

We provide hierarchical evidence for the role of signal transduction in cSVD. Signal transduction corresponds to the network of molecular pathways that convert environmental, mechanical, or chemical signals into coordinated cellular actions. Specifically, beyond associations at the cluster level, replicating in the i-Share cohort, we found specific ps-PRS associations involving VEGF, PDGF, ephrin (EPH), and EPHA2 forward signaling, all involved in processes relevant to cSVD. VEGF controls vascular permeability and angiogenesis[70], PDGF regulates pericytes and BBB integrity[71], and EPH receptors are likely critical to development of tight junctions and BBB permeability[72–74]. Of particular note, *EPHA2 forward signaling* was enriched for WMH across the

lifespan and plasma EPHA2 and EPHB4 protein levels were recently found to be associated with WMH, cognition, stroke, and dementia[75], with some drug repositioning perspectives[76]. Furthermore, VEGF and PDGF are key signaling pathways involved in aortic distensibility, itself causally linked to WMH[77].

ECM structure and function has previously been highlighted as a primary biological process underlying cSVD in pathways enrichment analyses from both genomic and proteomic studies[12]. Here, although the ECM pathway cluster did not replicate beyond the UK Biobank, we found robust evidence for association of WMH with the NPP1 pathway comprised in this cluster, involving *ENPP1* and well-known cSVD genes *COL4A1/2*[12,78]. This pathway was enriched in young adults and uniquely associated with increased risk of incident ischemic and, at nominal significance, hemorrhagic stroke and AD. *ENPP1* regulates vascular calcification, a hallmark of coronary artery disease, through production of inorganic pyrophosphate[79], as well as regulation of insulin receptor signaling and extracellular ATP levels.

In young adults, WMH may represent early stages of cSVD or early onset conditions like MS[80–83]. In this work, we showed that WMH global-PRS and 17 additional WMH ps-PRS were associated with WMH volume in young adults providing evidence that early-life WMH are consistent with white matter pathology found in older adults. In contrast, most ps-PRS enriched for WMH were distinct from those previously identified for MS[27], in line with another genetic study showing that cSVD and MS largely do not have shared genetic underpinnings[80].

In relation to clinical complications of cSVD, we originally demonstrate WMH global-PRS is predictive of stroke (both IS and ICH) and VMD, but not AD. Interestingly however, several WMH ps-PRS were associated with AD, including, at the cluster level, immune response to viral infections and carbohydrate metabolism. Notably, infection-related pathways (parasite, viral) are associated with AD in population-based cohorts, supporting emerging evidence that they may uniquely contribute to the cSVD–AD link[84–86].

We provide additional evidence that several other WMH ps-PRS may be of particular importance in older adults. *Elastic fibers,* a key component of the arterial ECM[87] and involved in atherosclerosis[88], were enriched in older community-persons and memory-clinic patients. C*hylomicron clearance* was significantly associated with ICH in the UK Biobank, which is likely driven by the inclusion of *APOE* in this pathway[89]. Other notable pathways— *organelle biogenesis,*

RHO GTPases, PDGF signaling, sphingolipid metabolism—showed consistent associations with both stroke and dementia, reinforcing shared cSVD mechanisms in these conditions.

In the memory clinic setting, several WMH ps-PRS were specifically associated with risk of incident dementia, such as *sphingolipid metabolism*. Indeed, sphingolipids form a core structural component of myelin, and their metabolites are implicated in neurodegeneration and neuroinflammation[90]. In particular, ceramides are highly dysregulated in stroke and cSVD patients and were shown to be associated with WMH volume and circulating biomarkers associated with brain injury[57,91–94]. The *disorders of galactose metabolism*, strongly and significantly associated with AD, is also related to a subclass of sphingolipids—galactocerebrosides—that are essential for myelin structure and stability[95(p)]. Sphingolipids are already an emerging target for diagnosis of treatment of cerebrovascular diseases, including cSVD and AD[96,97]. We also found the *peroxisomes* pathway to be associated with VMD in the memory-clinic. Peroxisomes play a critical role in neuroinflammation by regulating degradation of very-long-chain fatty acids and the metabolism of reactive oxygen species[98,99]. Particularly, peroxisome proliferator-activated receptor-γ (PPAR- γ), expressed in endothelial and vascular smooth muscle cells, is being currently investigated as a therapeutic target against cSVD[100,101]. We propose that ps-PRS could have potential for prognostic enrichment in cSVD or AD clinical trials that target specific pathways, such as those involving sphingolipids or peroxisomes.

Our study's primary strength is the comprehensive use of pathway-specific PRS, which, to our knowledge, is the first time this approach has been applied to cSVD MRI-markers. Unlike standard pathway enrichment analyses in GWAS (e.g., MAGMA), the ps-PRS approach directly quantifies an individual's genetic susceptibility within specific biological pathways. This provides particular value for disentangling heterogeneous conditions such as cSVD and offers greater context for genetically driven processes linking global WMH-PRS to stroke and dementia. We also leverage genotyped and MRI data from multiple independent cohorts spanning the adult lifespan and clinical spectrum to robustly validate our findings and identify pathways implicated in cSVD across, or specific to, life stages. We additionally highlight an array of secondary analyses pertinent to understanding the link between WMH, hypertension, early indicators of white matter microstructure, and stroke and dementia.

Our study has several limitations. First, associations of ps-PRS is susceptible to confounding and pleiotropy, and some associations may reflect pleiotropic effects of genes within these pathways. Thus, ps-PRS findings should be interpreted with caution and ideally combined with complementary methods such as Mendelian randomization to answer etiological questions. Second, we acknowledge multiple testing burden with only few WMH ps-PRS being associated with clinical outcomes after multiple testing correction. Thus, we position these analyses as exploratory, focusing on nominally consistent signals across follow-up cohorts and analyses. Third, there are multiple approaches to construct ps-PRS. We used an agnostic approach relying on existing pathway databases, which we consider appropriate for a heterogeneous disease such as cSVD with largely unknown mechanisms. However, hypothesis-driven approaches, as applied in AD[19,20,102], may be more informative and could complement our strategy. Fourth, while these cohorts spanned a wide age range, they largely underrepresented midlife (ages 40–65) when vascular risk factors are most relevant. We were also limited on sample size in follow-up cohorts (3C, Memento, i-Share) that are also impacted by selection factors (e.g., greater neurodegenerative vs vascular burden in Memento; survival bias in 3C). Larger longitudinal cohorts with repeated MRI measures would help clarify pathways involved in cSVD progression. Additional limitations include variation in MRI protocols and WMH algorithms across studies (e.g., 3C-Dijon with 1.5T imaging), heterogeneous clinical outcome definitions and potential misclassification, particularly for algorithm-based stroke and dementia definitions in the UK Biobank[103,104]. Finally, findings are limited to European ancestry, reducing generalizability and potentially exacerbating inequalities[105]. As GWAS sample sizes increase in diverse populations, applying ps-PRS across ancestries will be critical and may reveal ancestry-specific pathways.

Overall, pathways enriched in younger adults may indicate early processes in cSVD development, whereas those found only in older or memory-clinic populations likely reflect later pathological stages predisposing to stroke or dementia. Pathways related to lipid metabolism, ciliogenesis, and signal transduction, newly described in this study, may underlie cSVD across the lifespan, highlighting the interplay between metabolic dysfunction, low-grade inflammation, and vascular risk factors such as hypertension and obesity in driving endothelial injury and BBB disruption. Our findings point to potential pathways to target therapeutically, including lipid subclasses (omega-3/6 fatty acids, sphingolipids, peroxisomes), NPP1 in the ECM, and tyrosine kinase signaling, particularly EPHA2. We also propose that primary cilia may act as key endothelial signaling hubs

responsive to hemodynamic stress. Ps-PRS also have potential for risk stratifying cSVD complications, where certain biological endotypes may predispose more to CAA-related risk of ICH, or risk of AD. Future studies should validate these mechanisms across ancestries and integrate multi-omics, especially metabolomics, to refine causal pathways. By capturing pathway-specific genetic susceptibility, ps-PRS can aid biomarker discovery, personalized prevention, and clinical trial stratification in cSVD.

In conclusion, in a multicohort setting we demonstrate the predictive utility of a global WMH-PRS for cSVD and describe novel pathway-specific PRS associated with WMH and its clinical complications across the lifespan. These findings provide novel insights into cSVD mechanisms, with potential implications for therapeutic development and personalized prevention.

Figures & Tables

Figure 1. Overall study design. (A) Workflow used to derive and optimize global WMH-PRS using UK Biobank MRI data and 10-fold cross-validation. (B) Procedure to identify enriched WMH-psPRS and perform follow-up analyses in independent cohorts. (C) Secondary analysis was used to further investigate clinical and lifespan effects of selected pathways.

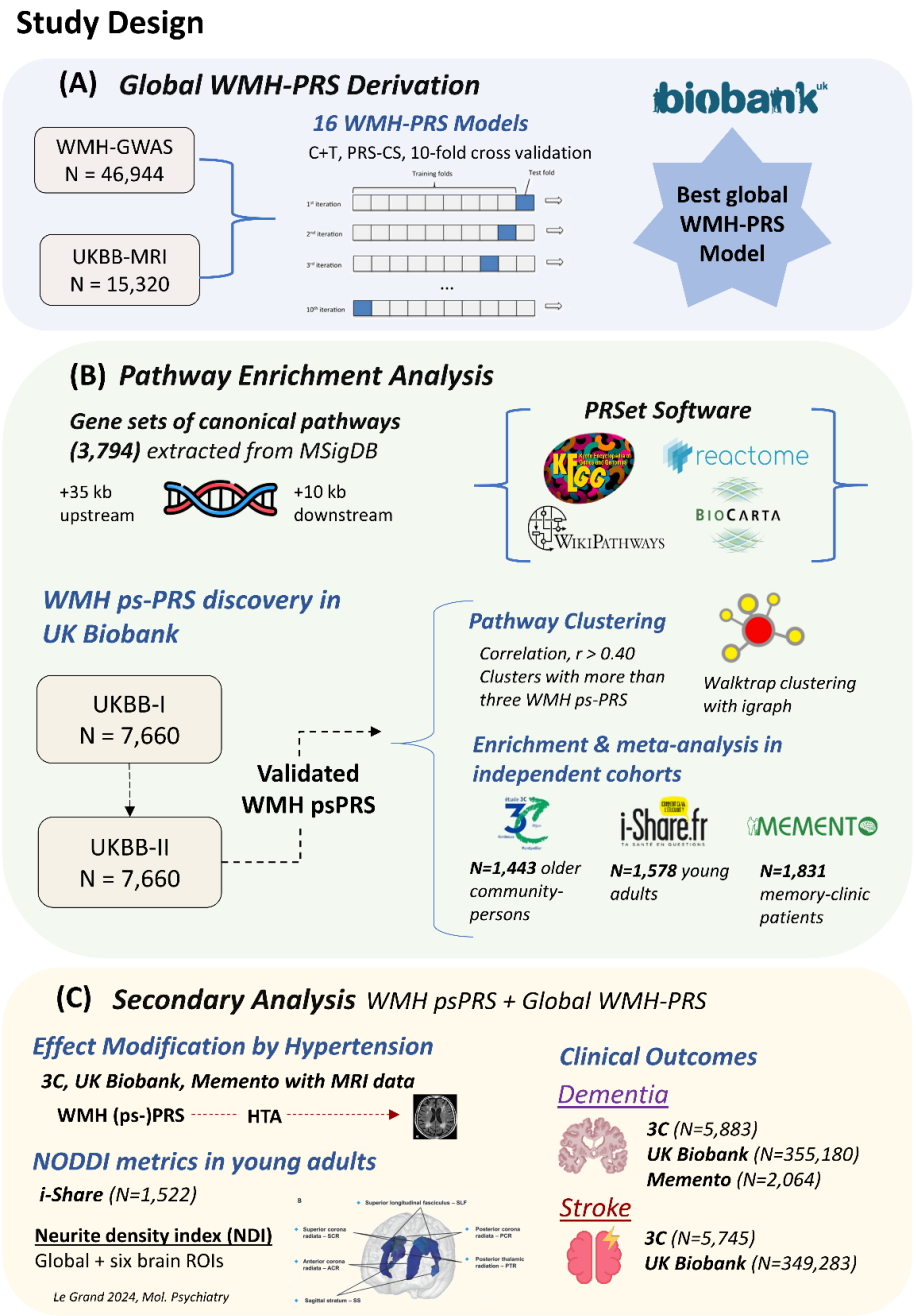
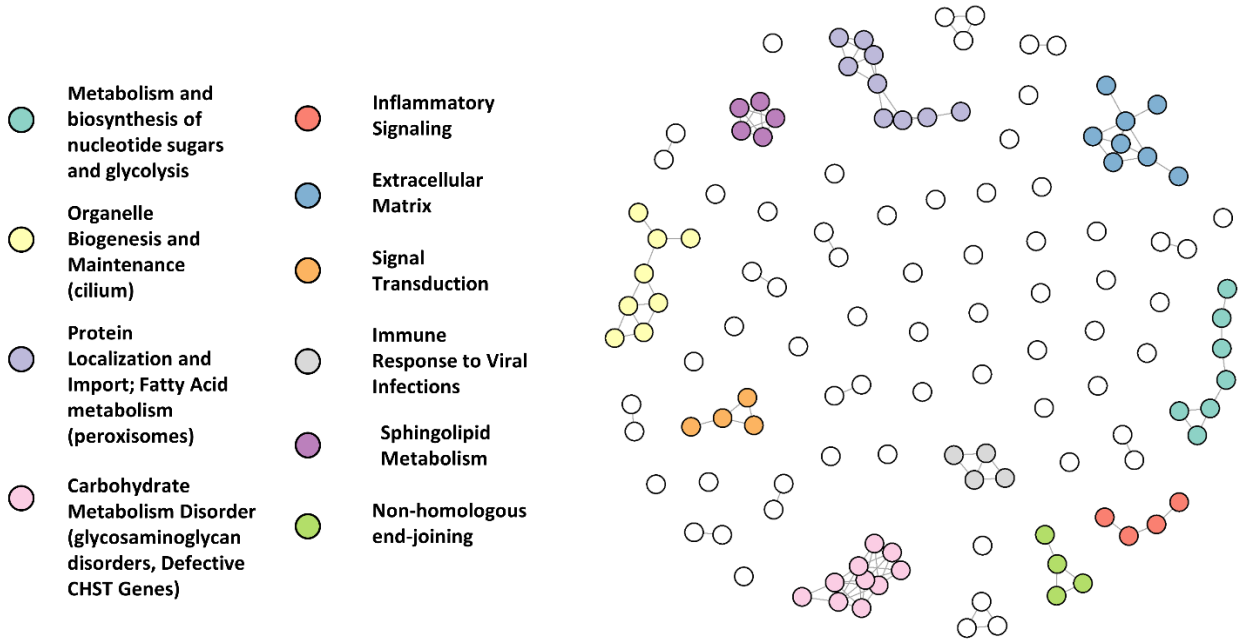


Table 1. Baseline characteristics of brain MRI studies.

	UKBB-I (N=7660)	UKBB-II (N=7660)	3C-Dijon (N=1443)	iSHARE (N=1578)	Memento (N=1831)
Age at MRI					
Mean (SD)	67.2 (7.16)	67.0 (7.17)	72.7 (4.12)	22.1 (2.26)	71.1 (8.52)
Median [Min, Max]	67.0 [52.0, 85.0]	67.0 [51.0, 85.0]	72.4 [65.2, 85.5]	21.7 [18.1, 34.9]	71.9 [32.9, 92.9]
Sex					
Female	4070 (53.1%)	4077 (53.2%)	869 (60.2%)	1131 (71.7%)	1138 (62.2%)
Male	3590 (46.9%)	3583 (46.8%)	574 (39.8%)	447 (28.3%)	693 (37.8%)
WMH Volume (cm3)					
Mean (SD)	5.57 (7.23)	5.65 (7.28)	5.57 (4.93)	1.31 (1.29)	9.51 (12.6)
Median [Min, Max]	3.17 [0.0990, 115]	3.23 [0.0930, 94.6]	4.07 [0.499, 50.0]	1.03 [0, 15.7]	5.44 [0.0800, 123]
Systolic Blood Pressure					
Mean (SD)	147 (21.1)	148 (20.8)	149 (22.8)	124 (12.6)	137 (18.1)
Missing	1871 (24.4%)	1882 (24.6%)	0 (0%)	713 (45.2%)	74 (4.0%)
Diastolic Blood Pressure					
Mean (SD)	79.6 (11.2)	79.9 (10.7)	84.9 (11.6)	76.1 (8.05)	77.3 (10.1)
Missing	1859 (24.3%)	1875 (24.5%)	0 (0%)	713 (45.2%)	74 (4.0%)
Anti-Hypertensive Medication					
Mean (SD)	0.272 (0.445)	0.262 (0.440)	0.426 (0.495)	0 (0)	0.333 (0.471)
Missing	95 (1.2%)	107 (1.4%)	0 (0%)	713 (45.2%)	74 (4.0%)
Hypertension					
Non-Hypertensive	1735 (22.7%)	1679 (21.9%)	626 (43.4%)	840 (53.2%)	1077 (58.8%)
Hypertensive	4530 (59.1%)	4576 (59.7%)	817 (56.6%)	25 (1.6%)	680 (37.1%)
Missing	1395 (18.2%)	1405 (18.3%)	0 (0%)	713 (45.2%)	74 (4.0%)

Figure 2. Clustering of the 127 enriched WMH ps-PRS in the UK Biobank. Clusters of more than 3 ps-PRS were defined based on Gene Ontology terms. Cluster-based ps-PRS were carried forward in follow-up analysis.



Cluster Name	Pathways
Metabolism and biosynthesis of nucleotide sugars and glycolysis	K AMINO SUGAR AND NUCLEOTIDE SUGAR METABOLISM; K GALACTOSE METABOLISM; K MEDI REFERENCE PINK PARKIN MEDIATED AUTOPHAGOSOME FORMATION; WP AEROBIC GLYCOLYSIS; K MEDI VARIANT MUTATION INACTIVATED OPTN TO PINK PARKIN MEDIATED AUTOPHAGOSOME FORMATION; K MEDI VARIANT MUTATION INACTIVATED P62 TO PINK PARKIN MEDIATED AUTOPHAGOSOME FORMATION; WP GLYCOLYSIS AND GLUCONEOGENESIS
Organelle Biosynthesis and Maintenance (cilium)	WP GENES RELATED TO PRIMARY CILIUM DEVELOPMENT BASED ON CRISPR; WP HEDGEHOG SIGNALING PATHWAY WP47; R ANCHORING OF THE BASAL BODY TO THE PLASMA MEMBRANE; R CILIUM ASSEMBLY; R ORGANELLE BIOGENESIS AND MAINTENANCE; WP BARDET BIEDL SYNDROME; WP HEDGEHOG SIGNALING PATHWAY WP4249; R RUNX2 REGULATES OSTEOBLAST DIFFERENTIATION
Protein Localization and Import and Fatty Acid metabolism (peroxisomes)	K ALPHA LINOLENIC ACID METABOLISM; WP OMEGA 3 OMEGA 6 FATTY ACID SYNTHESIS; WP EICOSANOID METABOLISM VIA CYCLOOXYGENASES COX; R PEROXISOMAL LIPID METABOLISM; R PEROXISOMAL PROTEIN IMPORT; R PROTEIN LOCALIZATION; R ALPHA LINOLENIC OMEGA3 AND LINOLEIC OMEGA6 ACID METABOLISM; K PEROXISOME; WP EICOSANOID METABOLISM VIA LIPOOXYGENASES LOX
Inflammatory Signaling	R COENZYME A BIOSYNTHESIS; K MEDI REFERENCE CA2 PLCD ITPR SIGNALING PATHWAY; R VITAMIN B5 PANTOTHENATE METABOLISM; WP RESISTIN AS A REGULATOR OF INFLAMMATION
Extracellular Matrix	R EXTRACELLULAR MATRIX ORGANIZATION; WP MIRNA TARGETS IN ECM AND MEMBRANE RECEPTORS; PID AVB3 INTEGRIN PATHWAY; R LAMININ INTERACTIONS; R SIGNALING BY PDGF; K SMALL CELL LUNG CANCER; R ASSEMBLY OF COLLAGEN FIBRILS AND OTHER MULTIMERIC STRUCTURES; BC NPP1 PATIIWAY
Signal Transduction	R SIGNALING BY VEGF; R SIGNALING BY RECEPTOR TYROSINE KINASES; R FCGAMMA RECEPTOR FCGR DEPENDENT PHAGOCYTOSIS; R PARASITE INFECTION
Non-homologous end-joining	K NON HOMOLOGOUS END JOINING; R EARLY PHASE OF HIV LIFE CYCLE; K MEDI REFERENCE NON HOMOLOGOUS END JOINING; K MEDI REFERENCE V D J RECOMBINATION
Carbohydrate Metabolism Disorder (glycosaminoglycan disorders, Defective CHST Genes)	R DEFECTIVE CHST14 CAUSES EDS MUSCULOCONTRACTURAL TYPE; R DEFECTIVE CHST3 CAUSES SEDCJD; R DEFECTIVE CHSY1 CAUSES TPBS; R DERMATAN SULFATE BIOSYNTHESIS; PID TOLL ENDOGENOUS PATHWAY; R DEFECTIVE B4GALT7 CAUSES EDS PROGEROID TYPE; R DISEASES ASSOCIATED WITH GLYCOSAMINOGLYCAN METABOLISM; R METABOLISM OF CARBOHYDRATES; R CS DS DEGRADATION; R A TETRASACCHARIDE LINKER SEQUENCE IS REQUIRED FOR GAG SYNTHESIS
Immune Response to Viral Infections	R INFECTIOUS DISEASE; R HIV INFECTION; R HIV LIFE CYCLE; R VIRAL INFECTION PATHWAYS
Sphingolipid Metabolism	WP SPHINGOLIPID METABOLISM OVERVIEW; WP SPHINGOLIPID METABOLISM INTEGRATED PATHWAY; WP SPHINGOLIPID PATHWAY; WP BIOSYNTHESIS AND TURNOVER OF 1 DEOXY SPHINGOID BASES; K SPHINGOLIPID METABOLISM

K = KEGG; R = Reactome; K MEDI = Kegg medicus reference; WP = wikipathways; PID = pathway interaction database; BC = Biocarta

Figure 3. Enrichment of WMH ps-PRS in Follow-up Cohorts. Pathways are grouped based on their biological cluster, represented by the first pathway in each grouping. The last, larger group are all pathways that did not belong to any cluster.

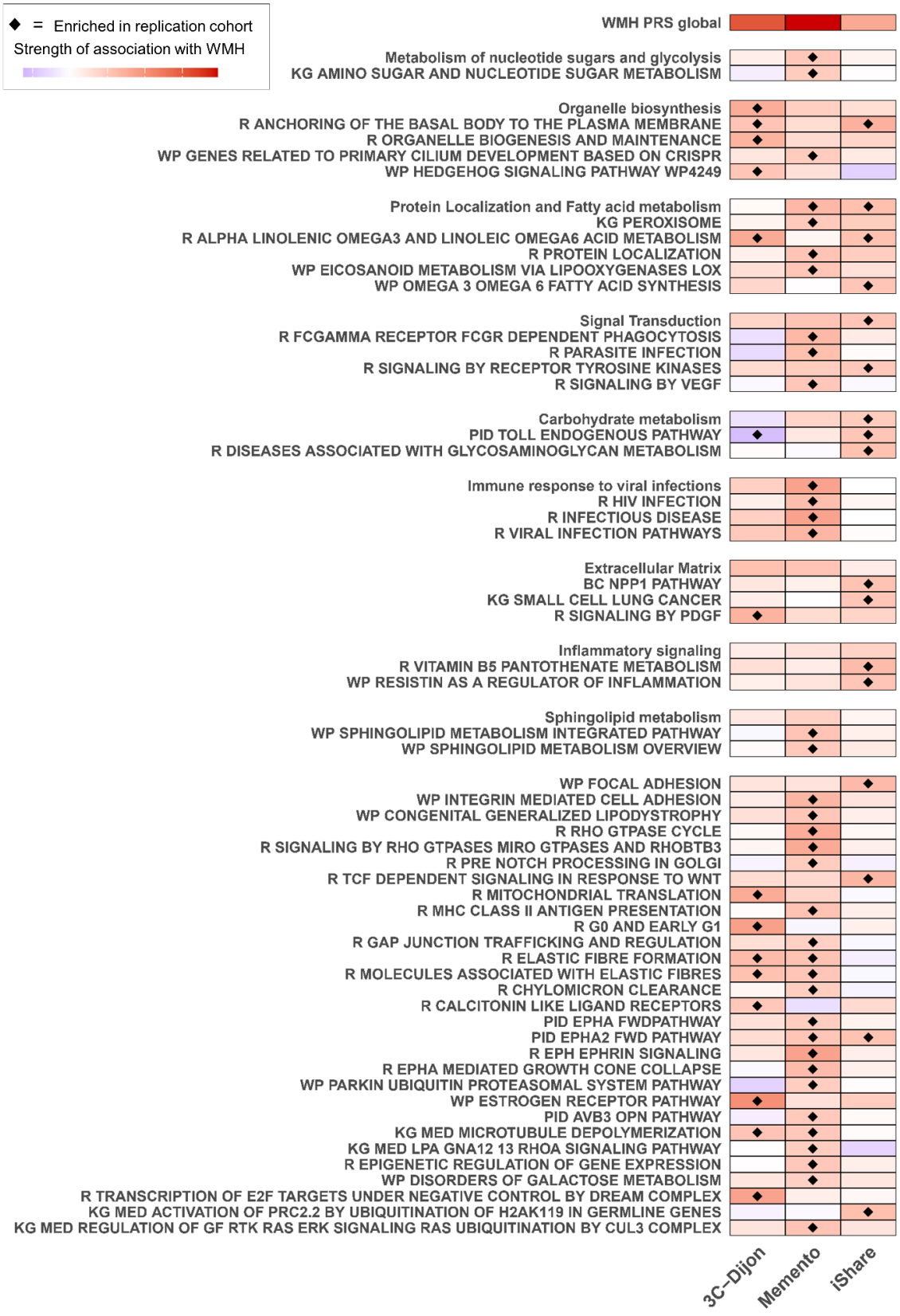
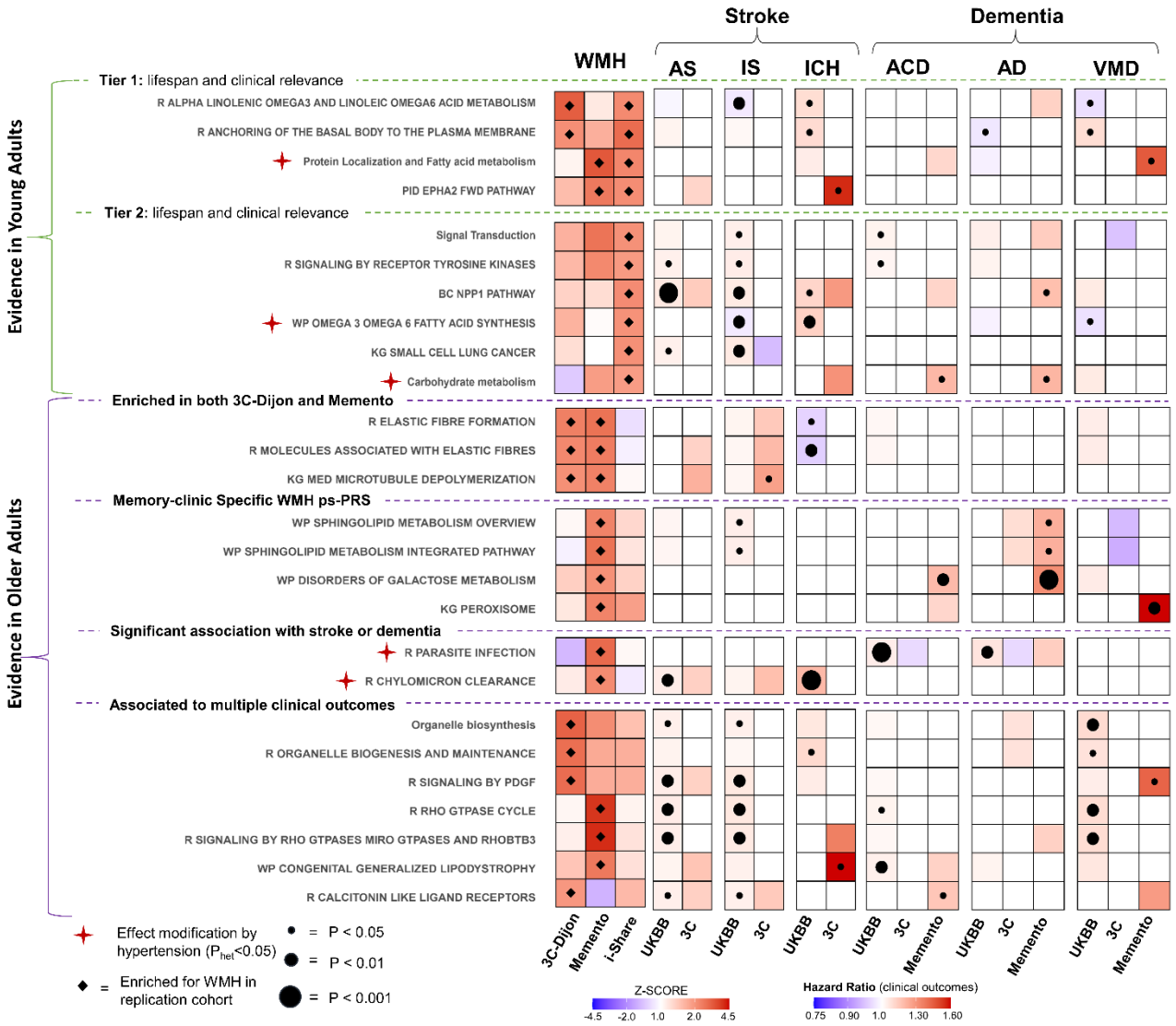


Figure 4. Prioritized findings from secondary analysis for association with clinical outcomes**.** We prioritized pathways in young adults based on lifespan and clinical relevance: (Tier 1) Enriched for WMH in follow-up cohorts of older and younger adults AND nominal association to stroke or dementia; (Tier 2) Enriched for WMH in i-Share only AND nominal association to stroke or dementia. In older adults we prioritize pathways with evidence for (1) Enrichment in both 3C and Memento; (2) Enriched for WMH and associated to clinical outcomes specifically in the memory-clinic; (3) strong association (P<0.001) to stroke or dementia; (4) association to multiple clinical outcomes.



References

1. White matter hyperintensities in vascular contributions to cognitive impairment and dementia (VCID): Knowledge gaps and opportunities. doi:10.1016/j.trci.2019.02.001

2. Wardlaw JM, Debette S, Jokinen H, et al. ESO Guideline on covert cerebral small vessel disease. *Eur Stroke J*. 2021;6(2):CXI-CLXII. doi:10.1177/23969873211012132

3. Backhouse EV, Boardman JP, Wardlaw JM. Cerebral Small Vessel Disease: Early-Life Antecedents and Long-Term Implications for the Brain, Aging, Stroke, and Dementia. *Hypertension*. 2024;81(1):54-74. doi:10.1161/HYPERTENSIONAHA.122.19940

4. Debette S, Markus HS. The clinical importance of white matter hyperintensities on brain magnetic resonance imaging: systematic review and meta-analysis. *BMJ*. Published online July 26, 2010. doi:10.1136/bmj.c3666

5. Prins ND, Scheltens P. White matter hyperintensities, cognitive impairment and dementia: an update. *Nat Rev Neurol*. 2015;11(3):157-165. doi:10.1038/nrneurol.2015.10

6. Tsuchida A, Boutinaud P, Verrecchia V, Tzourio C, Debette S, Joliot M. Early detection of white matter hyperintensities using SHIVA-WMH detector. *Hum Brain Mapp*. 2024;45(1):e26548. doi:10.1002/hbm.26548

7. Fang Y, Qin T, Liu W, et al. Cerebral Small-Vessel Disease and Risk of Incidence of Depression: A Meta-Analysis of Longitudinal Cohort Studies. *J Am Heart Assoc*. 2020;9(15):e016512. doi:10.1161/JAHA.120.016512

8. Direk N, Perez HS, Akoudad S, et al. Markers of cerebral small vessel disease and severity of depression in the general population. *Psychiatry Res Neuroimaging*. 2016;253:1-6. doi:10.1016/j.pscychresns.2016.05.002

9. van Sloten TT, Sigurdsson S, van Buchem MA, et al. Cerebral Small Vessel Disease and Association With Higher Incidence of Depressive Symptoms in a General Elderly Population: The AGES-Reykjavik Study. *Am J Psychiatry*. 2015;172(6):570-578. doi:10.1176/appi.ajp.2014.14050578

10. Garnier-Crussard A, Bougacha S, Wirth M, et al. White matter hyperintensities across the adult lifespan: relation to age, Aβ load, and cognition. *Alzheimers Res Ther*. 2020;12(1):127. doi:10.1186/s13195-020-00669-4

11. Kloppenborg RP, Nederkoorn PJ, Geerlings MI, van den Berg E. Presence and progression of white matter hyperintensities and cognition. *Neurology*. 2014;82(23):2127-2138. doi:10.1212/WNL.0000000000000505

12. Bordes C, Sargurupremraj M, Mishra A, Debette S. Genetics of common cerebral small vessel disease. *Nat Rev Neurol*. 2022;18(2):84-101. doi:10.1038/s41582-021-00592-8

13. Sargurupremraj M, Suzuki H, Jian X, et al. Cerebral small vessel disease genomics and its implications across the lifespan. *Nat Commun*. 2020;11(1):6285. doi:10.1038/s41467-020-19111-2

14. Persyn E, Hanscombe KB, Howson JMM, Lewis CM, Traylor M, Markus HS. Genome-wide association study of MRI markers of cerebral small vessel disease in 42,310 participants. *Nat Commun*. 2020;11(1):2175. doi:10.1038/s41467-020-15932-3

15. Duperron MG, Knol MJ, Le Grand Q, et al. Genomics of perivascular space burden unravels early mechanisms of cerebral small vessel disease. *Nat Med*. 2023;29(4):950-962. doi:10.1038/s41591-023-02268-w

16. Khera AV, Chaffin M, Aragam KG, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet*. 2018;50(9):1219-1224. doi:10.1038/s41588-018-0183-z

17. Le Grand Q, Tsuchida A, Koch A, et al. Diffusion imaging genomics provides novel insight into early mechanisms of cerebral small vessel disease. *Mol Psychiatry*. 2024;29(11):3567-3579. doi:10.1038/s41380-024-02604-7

18. Choi SW, García-González J, Ruan Y, et al. PRSet: Pathway-based polygenic risk score analyses and software. *PLOS Genet*. 2023;19(2):e1010624. doi:10.1371/journal.pgen.1010624

19. Darst BF, Koscik RL, Racine AM, et al. Pathway-Specific Polygenic Risk Scores as Predictors of Amyloid-β Deposition and Cognitive Function in a Sample at Increased Risk for Alzheimer's Disease. *J Alzheimer's Dis*. 2017;55(2):473-484. doi:10.3233/JAD-160195

20. Harrison JR, Foley SF, Baker E, et al. Pathway-specific polygenic scores for Alzheimer's disease are associated with changes in brain structure in younger and older adults. *Brain Commun*. 2023;5(5):fcad229. doi:10.1093/braincomms/fcad229

21. Bandres-Ciga S, Saez-Atienzar S, Kim JJ, et al. Large-scale pathway specific polygenic risk and transcriptomic community network analysis identifies novel functional pathways in Parkinson disease. *Acta Neuropathol (Berl)*. 2020;140(3):341-358. doi:10.1007/s00401-020-02181-3

22. Leeuw CA de, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLoS Comput Biol*. Published online 2015. doi:10.1371/journal.pcbi.1004219

23. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*. 2000;28(1):27-30. doi:10.1093/nar/28.1.27

24. Jassal B, Matthews L, Viteri G, et al. The reactome pathway knowledgebase. *Nucleic Acids Res*. 2020;48(D1):D498-D503. doi:10.1093/nar/gkz1031

25. Nishimura D. BioCarta. *Biotech Softw Internet Rep*. 2001;2(3):117-120. doi:10.1089/152791601750294344

26. Schaefer CF, Anthony K, Krupa S, et al. PID: the Pathway Interaction Database. *Nucleic Acids Res*. 2009;37(suppl_1):D674-D679. doi:10.1093/nar/gkn653

27. Shams H, Shao X, Santaniello A, et al. Polygenic risk score association with multiple sclerosis susceptibility and phenotype in Europeans. *Brain*. 2023;146(2):645-656. doi:10.1093/brain/awac092

28. Vascular Factors and Risk of Dementia: Design of the Three-City Study and Baseline Characteristics of the Study Population. *Neuroepidemiology*. 2003;22(6):316-325. doi:10.1159/000072920

29. Godin O, Dufouil C, Maillard P, et al. White Matter Lesions as a Predictor of Depression in the Elderly: The 3C-Dijon Study. *Biol Psychiatry*. 2008;63(7):663-669. doi:10.1016/j.biopsych.2007.09.006

30. Montagni I, Guichard E, Kurth T. Association of screen time with self-perceived attention problems and hyperactivity levels in French students: a cross-sectional study. *BMJ Open*. 2016;6(2):e009089. doi:10.1136/bmjopen-2015-009089

31. Le Grand Q, Satizabal CL, Sargurupremraj M, et al. Genomic Studies Across the Lifespan Point to Early Mechanisms Determining Subcortical Volumes. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 2022;7(6):616-628. doi:10.1016/j.bpsc.2021.10.011

32. Tsuchida A, Laurent A, Crivello F, et al. The MRi-Share database: brain imaging in a cross-sectional cohort of 1870 university students. *Brain Struct Funct*. 2021;226(7):2057-2085. doi:10.1007/s00429-021-02334-4

33. on behalf of the MEMENTO cohort Study Group, Dufouil C, Dubois B, et al. Cognitive and imaging markers in non-demented subjects attending a memory clinic: study design and baseline findings of the MEMENTO cohort. *Alzheimers Res Ther*. 2017;9(1):67. doi:10.1186/s13195-017-0288-0

34. Alfaro-Almagro F, Jenkinson M, Bangerter NK, et al. Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. *NeuroImage*. 2018;166:400-424. doi:10.1016/j.neuroimage.2017.10.034

35. Miller KL, Alfaro-Almagro F, Bangerter NK, et al. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat Neurosci*. 2016;19(11):1523-1536. doi:10.1038/nn.4393

36. Operto G, Chupin M, Batrancourt B, et al. CATI: A Large Distributed Infrastructure for the Neuroimaging of Cohorts. *Neuroinformatics*. 2016;14(3):253-264. doi:10.1007/s12021-016-9295-8

37. Maillard P, Delcroix N, Crivello F, et al. An automated procedure for the assessment of white matter hyperintensities by multispectral (T1, T2, PD) MRI and an evaluation of its between-centre reproducibility based on two large community databases. *Neuroradiology*. 2008;50(1):31-42. doi:10.1007/s00234-007-0312-3

38. Samaille T, Fillon L, Cuingnet R, et al. Contrast-Based Fully Automatic Segmentation of White Matter Hyperintensities: Method and Validation. *PLOS ONE*. 2012;7(11):e48953. doi:10.1371/journal.pone.0048953

39. Griffanti L, Zamboni G, Khan A, et al. BIANCA (Brain Intensity AbNormality Classification Algorithm): A new tool for automated segmentation of white matter hyperintensities. *NeuroImage*. 2016;141:191-205. doi:10.1016/j.neuroimage.2016.07.018

40. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203-209. doi:10.1038/s41586-018-0579-z

41. Auton A, Abecasis GR, Altshuler DM, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74. doi:10.1038/nature15393

42. Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data. *GigaScience*. 2019;8(7):giz082. doi:10.1093/gigascience/giz082

43. Ge T, Chen CY, Ni Y, Feng YCA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun*. 2019;10(1):1776. doi:10.1038/s41467-019-09718-5

44. Pain O, Glanville KP, Hagenaars SP, et al. Evaluation of polygenic prediction methodology within a reference-standardized framework. Plagnol V, ed. *PLOS Genet*. 2021;17(5):e1009021. doi:10.1371/journal.pgen.1009021

45. Kuhn M. Building Predictive Models in R Using the caret Package. *J Stat Softw*. 2008;28:1-26. doi:10.18637/jss.v028.i05

46. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*. 2011;27(12):1739-1740. doi:10.1093/bioinformatics/btr260

47. Csardi G, Nepusz T. The igraph software package for complex network research. *InterJournal*. 2006;Complex Systems:1695.

48. Csárdi G, Nepusz T, Müller K, et al. *Igraph for R: R Interface of the Igraph Library for Graph Theory and Network Analysis*. Zenodo; 2025. doi:10.5281/zenodo.14736815

49. Pons P, Latapy M. Computing communities in large networks using random walks. *arXiv:physics*. Published online 2005. doi:https://doi.org/10.48550/arXiv.physics/0512106

50. Reimand J, Kull M, Peterson H, Hansen J, Vilo J. g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res*. 2007;35(Web Server issue):W193-W200. doi:10.1093/nar/gkm226

51. Ashburner M, Ball CA, Blake JA, et al. Gene Ontology: tool for the unification of biology. *Nat Genet*. 2000;25(1):25-29. doi:10.1038/75556

52. The Gene Ontology Consortium, Aleksander SA, Balhoff J, et al. The Gene Ontology knowledgebase in 2023. *Genetics*. 2023;224(1):iyad031. doi:10.1093/genetics/iyad031

53. Zou Q, Wang M, Zhang D, Wei X, Li W. White Matter Hyperintensities in Young Patients from a Neurological Outpatient Clinic: Prevalence, Risk Factors, and Correlation with Enlarged Perivascular Spaces. *J Pers Med*. 2023;13(3):525. doi:10.3390/jpm13030525

54. Wardlaw JM, Smith C, Dichgans M. Small vessel disease: mechanisms and clinical implications. *Lancet Neurol*. 2019;18(7):684-696. doi:10.1016/S1474-4422(19)30079-1

55. Inoue Y, Shue F, Guojun B, Kanekiyo T. Pathophysiology and probable etiology of cerebral small vessel disease in vascular dementia and Alzheimer's disease.

56. Wang Z, Chen Q, Chen J, Yang N, Zheng K. Risk factors of cerebral small vessel disease: A systematic and meta-analysis. *Medicine*. 2021;100(51):e28229.

57. Harshfield EL, Sands CJ, Tuladhar AM, de Leeuw FE, Lewis MR, Markus HS. Metabolomic profiling in small vessel disease identifies multiple associations with disease severity. *Brain*. 2022;145(7):2461-2471. doi:10.1093/brain/awac041

58. Ye S, Feng K, Zeng G, et al. Frontiers | Deep and periventricular white matter hyperintensities exhibit differential metabolic profiles in arteriosclerotic cerebral small vessel disease: an untargeted metabolomics study. doi:10.3389/fnins.2025.1607242

59. Tiwari V, Simons M. Lipid metabolism and neuroinflammation: What is the link? *J Exp Med*. 2025;222(9):e20241232. doi:10.1084/jem.20241232

60. Xie Y, Liu S, Wang X, et al. Lipids, Apolipoproteins, Lipid-Lowering Drugs, and the Risk of Cerebral Small Vessel Disease: A Mendelian Randomization Study. *J Am Heart Assoc*. Published online August 20, 2024. doi:10.1161/JAHA.123.032409

61. Le Thuc O, García-Cáceres C. Obesity-induced inflammation: connecting the periphery to the brain. *Nat Metab*. 2024;6(7):1237-1252. doi:10.1038/s42255-024-01079-8

62. Gupta A, Thirugnanam K, Thamilarasan M, et al. Cilia proteins are biomarkers of altered flow in the vasculature. *JCI Insight*. 7(6):e151813. doi:10.1172/jci.insight.151813

63. Pala R, Jamal M, Alshammari Q, Nauli SM. The Roles of Primary Cilia in Cardiovascular Diseases. *Cells*. 2018;7(12):233. doi:10.3390/cells7120233

64. Luu VZ, Chowdhury B, Al-Omran M, Hess DA, Verma S. Role of endothelial primary cilia as fluid mechanosensors on vascular health. *Atherosclerosis*. 2018;275:196-204. doi:10.1016/j.atherosclerosis.2018.06.818

65. Mohieldin AM, Zubayer HSM, Omran AJA, et al. Vascular Endothelial Primary Cilia: Mechanosensation and Hypertension. Published online April 1, 2016. doi:10.2174/1573402111666150630140615

66. Ma N, Zhou J. Frontiers | Functions of Endothelial Cilia in the Regulation of Vascular Barriers. doi:10.3389/fcell.2020.00626

67. Mishra A, Duplaà C, Vojinovic D, et al. Gene-mapping study of extremes of cerebral small vessel disease reveals TRIM47 as a strong candidate. *Brain*. 2022;145(6):1992-2007. doi:10.1093/brain/awab432

68. Duplàa C, Delobel V, Boulestreau R, et al. Cerebral Small Vessel Disease genetic determinant TRIM47 controls brain homeostasis via the NRF2 antioxidant system. *bioRxiv*. Published online January 1, 2024:2024.10.08.616723. doi:10.1101/2024.10.08.616723

69. Gupta A, Thirugnanam K, Thamilarasan M, et al. Cilia proteins are biomarkers of altered flow in the vasculature. *JCI Insight*. 2022;7(6). doi:10.1172/jci.insight.151813

70. Lange C, Storkebaum E, de Almodóvar CR, Dewerchin M, Carmeliet P. Vascular endothelial growth factor: a neurovascular target in neurological diseases. *Nat Rev Neurol*. 2016;12(8):439-454. doi:10.1038/nrneurol.2016.88

71. Armulik A, Genové G, Mäe M, et al. Pericytes regulate the blood–brain barrier. *Nature*. 2010;468(7323):557-561. doi:10.1038/nature09522

72. Malik VA, Di Benedetto B. Frontiers | The Blood-Brain Barrier and the EphR/Ephrin System: Perspectives on a Link Between Neurovascular and Neuropsychiatric Disorders. doi:10.3389/fnmol.2018.00127

73. Coulthard MG, Morgan M, Woodruff TM, et al. Eph/Ephrin Signaling in Injury and Inflammation. *Am J Pathol*. 2012;181(5):1493-1503. doi:10.1016/j.ajpath.2012.06.043

74. Zhou N, Zhao WD, Liu DX, et al. Inactivation of EphA2 promotes tight junction formation and impairs angiogenesis in brain endothelial cells. *Microvasc Res*. 2011;82(2):113-121. doi:10.1016/j.mvr.2011.06.005

75. Sun Z, Harshfield EL, de Leeuw FE, et al. Proteins Involved in Endothelial Function and Inflammation Are Implicated in Cerebral Small Vessel Disease. *Stroke*. 2025;56(3):692-704. doi:10.1161/STROKEAHA.124.049079

76. Debette S, Caro I, Western D, et al. Proteogenomics in cerebrospinal fluid and plasma reveals new biological fingerprint of cerebral small vessel disease. *Res Sq*. Published online July 2, 2024:rs.3.rs-4535534. doi:10.21203/rs.3.rs-4535534/v1

77. Francis CM, Futschik ME, Huang J, et al. Genome-wide associations of aortic distensibility suggest causality for aortic aneurysms and brain white matter hyperintensities. *Nat Commun*. 2022;13(1):4505. doi:10.1038/s41467-022-32219-x

78. McNeilly S, Thomson CR, Gonzalez-Trueba L, et al. Collagen IV deficiency causes hypertrophic remodeling and endothelium-dependent hyperpolarization in small vessel disease with intracerebral hemorrhage. *eBioMedicine*. 2024;107:105315. doi:10.1016/j.ebiom.2024.105315

79. Nitschke Y, Hartmann S, Torsello G, et al. Expression of NPP1 is regulated during atheromatous plaque calcification. *J Cell Mol Med*. 2011;15(2):220-231. doi:10.1111/j.1582-4934.2009.00988.x

80. Brown RB, Traylor M, Burgess S, Sawcer S, Markus HS. Do Cerebral Small Vessel Disease and Multiple Sclerosis Share Common Mechanisms of White Matter Injury? *Stroke*. 2019;50(8):1968-1972. doi:10.1161/STROKEAHA.118.023649

81. Eikermann-Haerter K, Huang SY. White Matter Lesions in Migraine. *Am J Pathol*. 2021;191(11):1955-1962. doi:10.1016/j.ajpath.2021.02.007

82. Zhang LJ, Tian DC, Yang L, et al. White matter disease derived from vascular and demyelinating origins. *Stroke Vasc Neurol*. Published online September 12, 2023. doi:10.1136/svn-2023-002791

83. Zhang W, Cheng Z, Fu F, Zhan Z. Prevalence and clinical characteristics of white matter hyperintensities in Migraine: A meta-analysis. *NeuroImage Clin*. 2023;37:103312. doi:10.1016/j.nicl.2023.103312

84. Song J, Wang Y, Xie Z, Wei J, Wang J. Frontiers | Review of the mechanism of infection induced cerebral small vessel disease. doi:10.3389/fimmu.2025.1594891

85. Rippee-Brooks MD, Wu W, Dong J, Pappolla M, Fang X, Bao X. Viral Infections, Are They a Trigger and Risk Factor of Alzheimer's Disease? *Pathogens*. 2024;13(3):240. doi:10.3390/pathogens13030240

86. Letenneur L, Pérès K, Fleury H, et al. Seropositivity to Herpes Simplex Virus Antibodies and Risk of Alzheimer's Disease: A Population-Based Cohort Study. *PLOS ONE*. 2008;3(11):e3637. doi:10.1371/journal.pone.0003637

87. Halsey G, Sinha D, Dhital S, Wang X, Vyavahare N. Role of elastic fiber degradation in disease pathogenesis. *Biochim Biophys Acta BBA - Mol Basis Dis*. 2023;1869(5):166706. doi:10.1016/j.bbadis.2023.166706

88. Heinz A. Elastic fibers during aging and disease. *Ageing Res Rev*. 2021;66:101255. doi:10.1016/j.arr.2021.101255

89. Hostettler IC, Seiffge D, Wong A, et al. APOE and Cerebral Small Vessel Disease Markers in Patients With Intracerebral Hemorrhage. *Neurology*. 2022;99(12):e1290-e1298. doi:10.1212/WNL.0000000000200851

90. Ayub M, Jin HK, Bae J sung. Novelty of Sphingolipids in the Central Nervous System Physiology and Disease: Focusing on the Sphingolipid Hypothesis of Neuroinflammation and Neurodegeneration. *Int J Mol Sci*. 2021;22(14):7353. doi:10.3390/ijms22147353

91. Hannawi Y, Yanek LR, Kral BG, et al. White Matter Injury Is Associated with Reduced Manual Dexterity and Elevated Serum Ceramides in Subjects with Cerebral Small Vessel Disease. *Cerebrovasc Dis*. 2020;50(1):100-107. doi:10.1159/000511937

92. Ighodaro ET, Graff-Radford J, Syrjanen JA, et al. Associations Between Plasma Ceramides and Cerebral Microbleeds or Lacunes. *Arterioscler Thromb Vasc Biol*. 2020;40(11):2785-2793. doi:10.1161/ATVBAHA.120.314796

93. Mielke MM, Syrjanen JA, Bui HH, et al. Elevated Plasma Ceramides Are Associated With Higher White Matter Hyperintensity Volume—Brief Report. *Arterioscler Thromb Vasc Biol*. 2019;39(11):2431-2436. doi:10.1161/ATVBAHA.119.313099

94. Moseholm KF, Horn JW, Fitzpatrick AL, et al. Circulating sphingolipids and subclinical brain pathology: the cardiovascular health study. *Front Neurol*. 2024;15:1385623. doi:10.3389/fneur.2024.1385623

95. Barnes-Vélez JA, Yasar FBA, Hu J. Myelin lipid metabolism and its role in myelination and myelin maintenance. Accessed September 19, 2025. https://www.cell.com/the-innovation/abstract/S2666-6758(22)00156-4

96. Baloni P, Arnold M, Buitrago L, et al. Multi-Omic analyses characterize the ceramide/sphingomyelin pathway as a therapeutic target in Alzheimer's disease. *Commun Biol*. 2022;5(1):1074. doi:10.1038/s42003-022-04011-6

97. Custodia A, Romaus-Sanjurjo D, Aramburu-Núñez M, et al. Ceramide/Sphingosine 1-Phosphate Axis as a Key Target for Diagnosis and Treatment in Alzheimer's Disease and Other Neurodegenerative Diseases. *Int J Mol Sci*. 2022;23(15):8082. doi:10.3390/ijms23158082

98. Carmichael RE. Frontiers | Peroxisome dynamics and inter-organelle interactions in neuronal health and disease. doi:10.3389/fnmol.2025.1603632

99. Sarkar C, Lipinski MM. Role and Function of Peroxisomes in Neuroinflammation. *Cells*. 2024;13(19):1655. doi:10.3390/cells13191655

100. Koohi F, Harshfield EL, Gill D, Ge W, Burgess S, Markus HS. Optimizing treatment of cardiovascular risk factors in cerebral small vessel disease using genetics. *Brain*. 2025;148(6):1936-1949. doi:10.1093/brain/awae399

101. Lan LF, Zheng L, Yang X, Ji XT, Fan YH, Zeng JS. Peroxisome Proliferator-activated Receptor-γ Agonist Pioglitazone Ameliorates White Matter Lesion and Cognitive Impairment in Hypertensive Rats. *CNS Neurosci Ther*. 2015;21(5):410-416. doi:10.1111/cns.12374

102. Schork NJ, Elman JA. Pathway-Specific Polygenic Risk Scores Correlate with Clinical Status and Alzheimer's Disease-Related Biomarkers. *J Alzheimer's Dis*. 2023;95(3):915-929. doi:10.3233/JAD-230548

103. Wilkinson T, Schnier C, Bush K, et al. Identifying dementia outcomes in UK Biobank: a validation study of primary care, hospital admissions and mortality data. *Eur J Epidemiol*. 2019;34(6):557-565. doi:10.1007/s10654-019-00499-1

104. Rannikmäe K, Ngoh K, Bush K, et al. Accuracy of identifying incident stroke cases from linked health care data in UK Biobank. *Neurology*. 2020;95(6):e697-e707. doi:10.1212/WNL.0000000000009924

105. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Current clinical use of polygenic scores will risk exacerbating health disparities. *Nat Genet*. 2019;51(4):584-591. doi:10.1038/s41588-019-0379-x

Chapter 7 – Discussion

7.1 Summary of Findings

This thesis aimed to advance genetic risk prediction for cerebrovascular diseases—stroke, dementia, and cSVD—by refining and evaluating PRS. We focused on developing and applying innovative strategies to better capture disease heterogeneity and identify the populations and clinical settings most likely to benefit from PRS implementation. Our analyses utilized a diverse collection of individual-level clinical and genetic data encompassing a wide age range and clinical spectrum, sourced from large population-based cohorts, multi-ancestry biobanks, a memory clinic cohort, and a clinical trial of high-risk ICH survivors with AF. These studies featured international representation, including several French cohorts (3C, i-Share, Memento) participating in the RHU-SHIVA and IHU-VBHI initiatives, the UKBB, the AoU research program in the United States, and the European PRESTIGE-AF clinical trial. We also leveraged an extensive repository of GWAS summary statistics, most of which were publicly available, to construct PRS models for stroke, dementia, cSVD, and other related vascular and neurodegenerative traits. Importantly, we aimed to validate existing PRS models in clinically relevant prevention contexts and to adapt state-of-the-art PRS methodologies informed by advances in other polygenic diseases. This included integrating PRS across multiple related traits and decomposing genetic risk into biologically meaningful, pathway-specific components. Together, these efforts provide novel insights into the utility of PRS for primary and secondary prevention of stroke and dementia, complemented by a unique pathway-based approach that elucidates biological mechanisms underlying cSVD heterogeneity.

The first article aimed to enhance risk prediction of all-cause dementia by employing an integrative PRS framework (iPRS-DEM). This framework combines various dementia-related vascular and neurodegenerative traits and adapts previous approaches used for stroke to a competing risk setting (1–3). iPRS-DEM was associated with all-cause dementia independently of APOE in elderly community members (3C) and in a multi-ancestry biobank (All of Us), showing associations among Europeans and East Asians. We also demonstrated a strong association of iPRS-DEM in the memory clinic (Memento). Furthermore, iPRS-DEM informed risk stratification alongside APOE, further differentiating genetic risk among APOE-ε4 carriers and non-carriers. Thus, iPRS-DEM may provide a more comprehensive assessment of residual genetic risk beyond APOE that predisposes individuals to both neurodegenerative and vascular pathology. We emphasize the

strong predictive capability of iPRS-DEM in the memory clinic, a crucial real-world setting for dementia prevention. In dementia, PRS-based risk stratification offered a more precise estimate of genetic risk beyond APOE, facilitating the identification of high-risk individuals for primary prevention programs targeting lifestyle and vascular risk factors or for selection into clinical trials.

In the second article, we shifted to a secondary stroke prevention setting, investigating PRS in the PRESTIGE-AF trial of ICH survivors with AF at high risk of both ischemic and bleeding events. Despite the small sample size, we demonstrated the potential of an AF-PRS to differentiate ischemic from hemorrhagic stroke risk in the context of anticoagulation, in a highly selected sample with important clinical equipoise. Our findings suggest that AF-PRS could help identify individuals most likely to benefit from or be harmed by treatment with DOACs providing preliminary evidence for the translatability of PRS in secondary stroke prevention.

Finally, in the third article, we comprehensively assessed PRS for WMH, the most studied MRI marker of cSVD. Our focus was on ps-PRS to enhance our understanding of cSVD etiology from a life course perspective, leveraging data from young adults (i-Share) and older adults in either the general population (UK Biobank, 3C) or presenting at a memory clinic (Memento). Our findings support existing evidence regarding the role of pathways related to the structure and function of the ECM and introduce new evidence for pathways primarily involved in lipid metabolism, ciliogenesis, and signal transduction. Identified WMH ps-PRS showed differential association with clinical complications such as ischemic stroke, ICH, and dementia subtypes. In addition to providing insights for important pathways to target therapeutically, our results could eventually be incorporated into a risk prediction framework that utilizes pathway-specific PRS to stratify complications of cSVD and inform treatment strategies.

7.2 Interpretation of Findings

7.2.1 An integrative approach shows ability to stratify dementia risk in the general population comparable to stroke

PRS has been proposed as a tool to identify high-risk individuals in the general population who would benefit from early intervention. Unlike most clinical risk factors, PRS provides a stable lifetime risk assessment for a disease from birth and may capture risk independent of, and complementary to, existing clinical risk scores. However, the incremental benefit of adding stroke PRS to clinical risk factors is generally less evident despite associations with stroke independent of clinical risk factors (22,211,335). An integrative PRS approach, including multiple stroke subtypes and vascular risk factors, may capture some of the disease heterogeneity and has shown improved performance (10,211). A similar approach for ICH was also promising (212). In this thesis, we attempted to extend and adapt this approach to dementia, where we explicitly incorporate vascular genetic risk that likely contributes to mixed dementia pathology and account for the competing risk of death.

Dementia notably has a unique genetic architecture with *APOE* conveying an outsized amount of genetic risk alongside additional common risk alleles with very small effects. In the overall population, one study found that *APOE* combined with an AD-PRS, excluding *APOE*, achieved an AUC of 0.72-0.76, which dropped to 0.65 among non-APOE-ε4 carriers (336). Furthermore, a basic AD-GRS showed only modest improvement over *APOE* in population-based cohorts (12). Most prior studies focused on case-control data and AD-specific outcomes, whereas, due to diagnostic challenges, an all-cause dementia outcome is more practical for primary prevention in the general population. We demonstrate that iPRS-DEM, an integrative PRS encompassing polygenic risk of AD, neurodegenerative endophenotypes, as well as vascular brain disease (stroke, cSVD) and vascular risk factors, is associated with all-cause dementia in community-based older adults (>65 years). We additionally showed that it may enhance risk stratification more than overall risk prediction. Indeed, the prediction of APOE + iPRS-DEM was similar to previous studies ($AUC_{10\text{-}year}$=0.741), with only a marginal improvement when adding iPRS-DEM over APOE *($AUC_{10\text{-}year}$=+0.002)*. Individuals in the top 6% of iPRS-DEM among non-APOE-ε4 carriers have a 2-fold increased relative risk and a cumulative risk of 13.2% [6-20.4%] by age 80, comparable to the top 44% among APOE-ε4 carriers, who have similar relative-risk and a cumulative risk of 12.4% [7.4-17.4%] by age 80. These findings provide important evidence for the utility of PRS-based risk stratification for dementia beyond APOE in the general population. Additionally, to my knowledge, most studies have not investigated dementia PRS in the context

of established risk factors. One recent study examined a 68-SNP AD-GRS in relation to the Lifestyle for Brain Health risk score (LIBRA) within the 3C study and found evidence for additive effects on dementia risk, although it did not assess prediction (337). Similar to findings for stroke, prediction performance in the general population is comparable to that of established risk factors.

We also made an initial attempt to externally validate iPRS-DEM across diverse ancestries using AoU data, and show transportability in East Asians. The GIGASTROKE project indicated that an iPRS for stroke derived from European ancestry had reduced association and prediction performance in non-European ancestries, highlighting the benefits of an ancestry-specific iPRS in East Asians (10). Our results for dementia align with this, though we have limited power in All of Us to detect significant associations in African-Americans and Hispanics. Recent efforts have aimed to adapt standard AD-PRS to non-European ancestries, including East Asians, Indigenous Africans, African-Americans, and admixed populations in Latin America (338–340). Generally, AD-PRSs, similar to stroke, maintain significant but diminished predictive associations with AD risk in most non-European populations. The evidence is promising that PRS prediction for cerebrovascular diseases in non-European populations is feasible, although substantial gaps in prediction performance remain.

7.2.1.1 PRS for cerebrovascular disease is not yet ready for use in the general population

Interest in using PRS for primary prevention is particularly strong for diseases with established risk thresholds, population-based screening programs and effective health management strategies. Breast cancer serves as a prime example, as PRS improves prediction beyond highly penetrant genes like *BRCA1* and *BRCA2*, and can meaningfully improve the efficiency of established screening programs (341,342). Similarly, the European Society of Cardiology has established guidelines for the primary prevention of cardiovascular disease that incorporate clinical risk scores into risk assessments (**Figure 20**) (324). Although PRS are not yet recommended, there is growing evidence supporting their value when integrated into clinical risk scores to guide statin initiation (22,325). In fact, PRS may be most relevant for

Table 5 Cardiovascular disease risk categories based on SCORE2 and SCORE2-OP in apparently healthy people according to age

	<50 years	50–69 years	≥70 years[a]
Low-to-moderate CVD risk: risk factor treatment generally not recommended	<2.5%	<5%	<7.5%
High CVD risk: risk factor treatment should be considered	2.5 to <7.5%	5 to <10%	7.5 to <15%
Very high CVD risk: risk factor treatment generally recommended[a]	≥7.5%	≥10%	≥15%

Figure 20. From the 2021 *ESC* guidelines. Here 10-year risk estimates from a clinical risk score, SCORE2, are used to inform initiation of treatment. Treatment for CVD typically includes statin initiation to control LDL cholesterol.

individuals near the threshold for risk initiation, as they can adjust risk levels either upward or downward. One study estimated the net reclassification index (NRI) of 10-year risk—which indicates how many people change risk categories—at 5.9% (4.7-7.0%) when including a CAD-PRS. This increase was particularly pronounced in younger subgroups, reaching 15.4% (11.6-19.3%) (22,343). Consequently, guidelines are being developed to incorporate these scores into clinical practice (325), and randomized trials are currently exploring the benefits of including PRS risk assessments for both breast cancer and cardiovascular disease primary prevention (341,344,345). Such trials do not yet exist for the prevention of stroke or dementia to our knowledge.

Stroke risk likely is attributable, in part, to independent biological pathways, such as those involved in coagulation (21). Incorporating stroke PRS may enhance risk assessment for cardiovascular disease beyond what CAD-PRS can offer, particularly for intermediate-risk groups (360). One study modeled a screening scenario using both ischemic stroke PRS and CAD-PRS to reclassify individuals with a 10-year cardiovascular disease risk of 5-10% based on conventional risk factors (346). Their findings indicated that this targeted screening approach could identify one additional cardiovascular disease event for every 340 individuals screened, making it 15 times more effective than a general PRS screening strategy (360). Overall, the specific applications of PRS in stroke prevention in healthy populations remain unclear.

In healthy populations, dementia risk assessment focuses on identifying and managing modifiable risk factors, and genetic testing is not standard in clinical practice (129,347). Even established clinical risk scores, such as the CAIDE (Cardiovascular Risk Factors, Aging, and Incidence of Dementia), which can include *APOE*, have limited effectiveness for personalized dementia risk assessment in the general population (348). Recently, a machine-learning model that incorporated *APOE*-ε4 along with nine other predictors showed significantly improved prediction in the UK Biobank (349). However, it is evident that we are still in need of evidence-based guidelines regarding integration of dementia risk prediction into routine primary care, and more evidence is needed evaluating PRS and other biomarkers in these models (142). There is also considerable interest in the potential of dementia screening programs. A trial examining dementia screening in older adults (>65) within primary care found neither harm nor benefit from the intervention (350). Alternative dementia screening strategies that utilize biomarkers, genetics, or simple mobile applications (e.g., for CAIDE) have been proposed to assist in diagnoses or detect early preclinical signs (351,352). Overall, our findings align with the prevailing view that PRS are not yet sufficiently robust for inclusion in individual risk prediction for dementia screening or other

applications (142,336). Nevertheless, as new screening and risk assessment strategies develop, genetic testing both with *APOE* and PRS should be carefully considered and evaluated. Their utility will ultimately largely depend on the preventive strategies that can be proposed.

Compared to diseases such as breast cancer and CAD, there is limited evidence and no standardized guidelines for adressing the clinical risk of stroke and dementia in healthy populations. Current primary prevention strategies largely mirror those for cardiovascular disease, focusing on managing vascular risk factors with prescription medications or lifestyle changes. Multi-modal interventions specifically tailored to dementia have been proposed (129). A much larger evidence base is needed for the successful adoption and implementation of PRS in primary prevention to assess the risk of cerebrovascular disease in the general population. Nevertheless, our findings contribute to the existing evidence suggesting that genetic risk factors beyond APOE are likely to enhance risk prediction and stratification. Additionally, genetic risk assessments for dementia should consider vascular pathology which would be in alignment with both the focus on modifiable vascular risk factors, and the potential to target VCI in dementia prevention trials.

7.2.2 PRS for cerebrovascular disease likely provide greater utility in high-risk populations.

Populations at high-risk for cerebrovascular disease present a more actionable to target for PRS-based risk stratification due to more established clinical guidelines, and greater risk-benefit trade-offs. The findings of this work provide evidence of PRS utility in the memory clinic, and for secondary stroke prevention among ICH survivors with AF.

7.2.2.1 PRS could be integrated in future second-generation memory clinics

Memory clinics are traditionally outpatient facilities focused on diagnosis, staging, treatment, and rehabilitation for cognitively impaired individuals – from those with subjective cognitive complaints to severe dementia. They represent a pragmatic starting point for dementia risk assessment and prevention. Recently, a new model called Brain Health Services (BHS) has been proposed as second-generation memory clinics whose primary mission is evidence-based prevention focused on cognitively unimpaired individuals who are potentially at risk (142). In BHS, PRS are viewed as a scalable and central component to comprehensive genetic risk assessment alongside *APOE*. Presently, PRS studies provide evidence for the ability of standard AD-PRS (mostly comprised of GWS variants, excluding *APOE*) to predict MCI conversion to dementia (226,339,353,354), and a few show associations of AD-PRS directly in a memory-clinic setting (12,355,356). Generally, PRS studies in individuals with MCI focus on AD-type dementia and show mostly consistent findings that AD-PRS can predict MCI conversion to AD

(226,339,353,354), although some observed stronger associations among non-*APOE-ε4* carriers (339,354) or attenuated risk for those with low AD-PRS among *APOE-ε4* carriers (356). Yet, only one study investigated all-cause dementia and did not find any association, even reporting inverse associations with non-AD dementia (356).

The work presented in Article 1 provides valuable evidence regarding the utility of PRS in memory clinics. Unlike other PRS studies that have primarily included only age, sex, and sometimes education as risk factors, we also investigated a set of vascular risk factors. Our study found that only education and family history of dementia are predictive of all-cause dementia, while iPRS-DEM and *APOE* exhibit significantly greater predictive ability than these factors. In memory clinics, genetic risk may contribute more to overall risk than established factors when compared to the general population, though this needs further validation. Consistent with other studies, we emphasize the potential benefits of defining genetic risk groups, as iPRS-DEM and *APOE-ε4* can identify individuals in the highest genetic risk category, facing a 5 to 7-fold greater risk than those in the lowest risk group (33,34). Additionally, *APOE-ε4* carriers with low PRS may experience reduced risk and unnecessary intervention may be avoided (356). In Article 3, we also provide initial evidence that some WMH ps-PRS are uniquely association with dementia in the memory clinic, highlighting specifically sphingolipid and peroxisome pathways. Both sphingolipids and peroxisome related pathways have been suggested as therapeutic targets for both cSVD and AD (357,358).

Our findings combined with existing evidence highlight that PRS could refine risk assessment beyond *APOE* and demographic factors, which would be particularly important among cognitively normal individuals with subjective cognitive complaints. Holistically, knowledge of genetic risk informs clinicians on the necessity and intensity of risk reduction strategies and would motivate initial multi-modal interventions focusing on lifestyle, fitness, nutrition, and management of vascular risk factors. Evidence shows multi-modal interventions may be particularly effective in those at highest genetic risk (140,142). High risk patients could also be prioritized for diagnostic workup with more precise imaging and circulating biomarkers, and/or recruited into clinical trials to improve their efficiency. Given iPRS-DEM incorporates vascular risk factors, it could uniquely identify individuals for trials targeting therapies not specific to AD biomarkers like amyloid or tau. Disease-modifying therapies, like anti-amyloid immunotherapies, are available but come with potential of micro-hemorrhages is the form of amyloid-related imaging abnormalities (ARIA) for which *APOE-ε4* and WMH are a major risk factor likely through CAA-related risk (359,360).

iPRS-DEM and relevant WMH ps-PRS, which may also predispose to CAA, should be investigated to inform risk of these novel therapies.

In summary, genetic testing with *APOE* and dementia PRS has potential benefit in a memory-clinic setting and can help guide more actionable and pragmatic dementia prevention than in primary care. Currently, genetic testing focuses on identifying causative, monogenic forms and determining *APOE* genotype to assess treatment safety for amyloid- modifying therapies in AD (347,361,362). There is additional evidence that more-relaxed eligibility criteria and screening for monogenic disorders significantly improves the detection of individuals carrying pathogenic mutations in a memory clinic (362). Eventually, new guidelines will be needed to incorporate *APOE* and dementia PRS into genetic risk assessment in future memory-clinics following the BHS framework. Before they can be implemented, PRS require some additional evidence for their predictive capacity in the context of other risk factors and novel biomarkers among those with MCI or cognitively normal individuals with subjective cognitive complaints. Importantly, we will additionally need structured risk communication protocols and established incremental risk thresholds (142). The findings presented in this thesis advance evidence for the potential utility in the memory clinic.

7.2.2.2 PRS has potential to inform anticoagulation therapy in secondary stroke prevention

In guidelines for AF management, long-term oral anticoagulation therapy is based on stroke risk profile using the CHA_2DS_2-VASc score balanced against bleeding risk (usually measured with the HAS-BLED score) (363). Among patients with high risk of both bleeding and ischemic events, like ICH survivors with AF in the PRESTIGE-AF trial, treatment guidelines are uncertain. The PRESTIGE-AF trial provided inconclusive results for efficacy and safety of DOACs in this population. However, existing evidence that PRS may be able to stratify risk and help guide anticoagulation therapy in this setting motivated the genetic sub-study in Article 2 of this thesis.

Prior PRS studies consistently demonstrated that PRS for ischemic stroke or AF are independent predictors of ischemic stroke in patients with AF or more general cardiometabolic disease (10,215,364,365). Specifically, both AF and ischemic stroke PRS were strongly associated with stroke risk among individuals with low CHA2DS2-VASc (<2), successfully identifying patients who possess a high genetic risk comparable to those with higher clinical risk scores (215,364). One study also showed an ICH-PRS can effectively predict bleeding events among patients on anticoagulants (mostly warfarin) in the UK Biobank, and has added value when integrated into HAS-BLED (213). In Article 2, we investigated the ability of ICH-PRS, the IS-iPRS from the

Gigastroke project, the AF-PRS from the latest GWAS published earlier this year, and a WMH-GRS to predict stroke risk in PRESTIGE-AF. Our findings provide initial evidence for the utility of PRS in this population, most directly through being able to assess meaningful DOAC risk reduction among those with high AF-PRS who experienced more ischemic strokes and no bleeding events. Our main limitation here was sample size, and we emphasize the need to replicate these findings in future clinical trials, such as the ASPIRE-AF trial in the US. We underscore the value of post-hoc analysis of clinical trials to evaluate PRS, especially when evidence is inconclusive as was the case in PRESTIGE-AF.

The findings here also support that PRS could be useful to predict treatment response for DOACs as seen with statins and PCSK-9 inhibitors to prevent CAD (232,233,242). However, more investigations are needed to see if genetic risk modulates DOAC effectiveness generally. There is also potential for PRS that integrate pharmacogenomic variants, for instance those directly implicated in DOAC metabolism and anticoagulation-related bleeding risk (366,367), which have could refine PRS for predicting treatment response (368).

When it comes to the immediate application of PRS, they are likely most valuable to guide clinical decisions in high-stakes scenarios of clinical equipoise. Stroke survivors, often presenting with multiple risk factors and comorbidities, are a unique patient group with high absolute risk for adverse outcomes, including stroke recurrence, disability, and dementia, making them a key target for implementing advanced risk stratification and prevention strategies (369). Ideally, these patients would have the most comprehensive risk assessment, and clinicians should have all validated imaging, biomarker and genetic risk tools at their disposal. Given these patients are highly selected and already likely at high genetic risk, it is unlikely that PRS trained and validated in population-based studies are overly transferable. Indeed, genetic variants that predict the index event show inconsistent associations with recurrent events (370). Therefore, it would likely be beneficial if SNP effects could be re-calibrated for these populations. Large-scale biobanks could be efficient training data sources – where proxy datasets representing high-risk populations can be derived.

PRS has the potential to refine individual risk assessment in high-risk patient populations. In addition to DOAC use in ICH survivors with AF, one can also suppose a similar risk stratification approach for individuals eligible for disease-modifying therapies for dementia (described above). This evidence also encourages that, in addition to PRS being generalizable across populations and for primary prevention, PRS should also be tailored to high-risk patient populations. Given the challenges at the general population level (e.g., lack of equity across ancestries), high-risk

scenarios and secondary prevention likely are more feasible and represent pragmatic starting points.

7.2.3 The beginning of translatable PRS for cSVD

Prior to this thesis, PRS for cSVD had been only sparsely explored with WMH being the only cSVD MRI-marker with sufficiently powered GWAS to derive PRS. Article 3 provides the first comprehensive assessment of the predictive ability of PRS derived from WMH GWAS. In terms of overall risk prediction, we highlight a global WMH-PRS could be a useful predictor in risk prediction models for cSVD complications like stroke and dementia. The major novelty of this work is the pathway-based PRS approach for its potential to identify potential pathways implicated in the disease. A major theme in the pathways identified was inflammatory response that could result from specific immune signaling pathways and altered lipid metabolism, particularly involving fatty acids. Additionally, its possible that primary cilia, by sensing shear stress due to blood pressure, could be key in linking hypertension to dysfunctional vascular endothelium in cSVD.

Recent investigations into cSVD etiology have leveraged a multi-omics approach, like proteomics (371,372), transcriptomics (373,374), and metabolomics (375). These studies have identified potential biological markers and signatures of cSVD, highlighting potential druggable targets, implicating the microbiota, and identifying mechanisms shared with stroke and dementia. Still, we lack a clear understanding of cSVD etiology, and there is now a need to integrate findings across omics layers. Further we need to identify which cSVD mechanisms are involved at different stages of cSVD progression and those that predispose to specific clinical outcomes. We provide evidence that some pathways are implicated in cSVD across the lifespan, and others that are enriched either earlier or later in disease progression, including those predisposing to stroke and dementia sub-types. The set of WMH ps-PRS identified presents an interesting starting point to contextualize multi-omics findings in cSVD. Especially given that cSVD lacks a precise clinical definition and is currently characterized by imaging markers, more convergence evidence is needed to pinpoint robust mechanisms. A possible future exploration could involve ps-PRS enriched by markers from metabolite or protein quantitative trait loci (mQTL, pQTL) (371,376). Integration of multi-omics information within a ps-PRS provide a potential avenue for future exploration, complementing existing approaches.

These WMH ps-PRS scores have potential to disentangle cSVD heterogeneity and define cSVD endotypes, for example, atherosclerotic cSVD, immune-related inflammatory cSVD, or CAA-driven cSVD. This could be particularly valuable in a risk-prediction setting by identifying genetic

profiles that predispose individuals to specific cSVD endotypes, each of which may confer risk for different downstream complications—for example, a greater propensity for lacunar stroke, CAA-related ICH, VCI, or even AD. However, before such applications are feasible, this strategy must be evaluated across a broader range of cSVD MRI markers that likely capture distinct underlying biological processes, as well as more deeply investigate how these pathways may be linked to specific clinical cSVD manifestations.

cSVD-dedicated clinics are just starting to emerge, and there is a pressing need for health services and guidelines dedicated to the prevention and management for cSVD (377). Individuals are most likely to be identified if they exhibit overt symptoms (e.g. stroke), or have suspected monogenic disorder (e.g, CADASIL) or CAA. Currently, we would miss most covert cSVD cases since neuroimaging assessment is not a routine part of daily practice. Barriers in cSVD management are due to a lack of overall awareness of cSVD stemming from heterogeneity of clinical symptoms, a lack of standardized diagnosis, and no established risk thresholds (377). PRS could have clinical utility to address these challenges in a few ways. First, PRS could be used to screen individuals for neuroimaging assessment for covert cSVD. Second, as we provide initial evidence for here, PRS can help distinguish heterogeneity in cSVD. Third, PRS could be used in individuals with suspected monogenic cSVD (e.g. CADASIL) to better characterize risk of severe complications. Indeed, there is evidence that PRS could modulate the penetrance of monogenic disorders. This has been shown in familial hypercholesterolemia where PRS for LDL-C or CAD can up- or -down grade the clinical severity and penetrance of a pathogenic variant (378). We are likely far away from direct clinical use of PRS in cSVD, but high-risk individuals are already being identified, and PRS eventually could be a valuable tool to assess risk in these patients.

7.3 Limitations and considerations to improve PRS risk prediction

While polygenic risk scores hold promise for improving disease risk stratification, a number of well-recognised limitations must be contextualised when interpreting the findings of this thesis. These span methodological challenges in PRS construction and generalisability, biases inherent to the GWAS study designs on which PRS depend, conceptual limitations of the polygenic-additive model, and practical barriers to clinical implementation — each discussed in turn in sections 7.3.1–7.3.5. Broadly, PRS remain probabilistic population-level tools whose individual-level precision is constrained by the quality and diversity of their training data, the assumptions of the underlying model, and the degree to which genetic risk can be disentangled from environmental and social context. These limitations are particularly consequential for cerebrovascular disease given its aetiological heterogeneity and the modest predictive ability of PRS for these conditions.

A condensed overview of the major limitations, their specific impact on the work presented in this thesis, and the steps taken to address them where possible are presented in **Table 5**. The subsequent subsections discuss each in greater depth.

Limitation	Impact on this work	How limitation is addressed
Ancestry bias in GWAS training data (§7.3.1)	iPRS-DEM was trained on predominantly European-ancestry GWAS; poor transferability demonstrated in Hispanic and African-American participants in All of Us	Attempt to validated iPRS-DEM in multi-ancestry All of Us cohort; applied cross-ancestry GWAS (GIGASTROKE); identified this as a key direction for future work
Lack of standardisation in PRS calculation and reporting (§7.3.2)	Variability between PRS methods can lead to discordant individual risk classifications; choice of algorithm influences model performance	Used the GenoPred reference-standardised framework across cohorts; deposited iPRS-DEM on PGS Catalog for independent validation
Sample overlap and study design biases (§7.3.3)	Inflation of association statistics observed in 3C/Memento due to overlap with base AD-GWAS; participation bias in biobank-derived GWAS may distort effect sizes; generalizability may be limited	Systematically re-ran GWAS meta-analyses excluding 3C and Memento; acknowledged participation bias as a potential source of residual confounding; Validated in multiple settings (young adults, memory-clinic, stroke survivors)
Limitations of the polygenic-additive model (§7.3.4)	PRS for stroke, dementia, and cSVD explained at most 2–3% of variance; model assumptions of homogeneity and additivity poorly suited to heterogeneous cerebrovascular disease	Developed integrative (iPRS-DEM), and pathway-informed PRS (for WMH) to partially capture disease heterogeneity; interpreted findings in context of known model limitations
Barriers to clinical implementation (§7.3.5)	No validated clinical pipeline exists for PRS in cerebrovascular disease; limited RCT evidence for PRS utility; stroke and dementia absent from current clinical implementation initiatives due to insufficient clinical validity	Findings framed as contributions to the research evidence base and do not promote widespread PRS use; integration with clinical risk factors discussed as prerequisite for utility; high-risk clinical cohort findings provide a potential translational bridge

Table 5. Summary of PRS limitations and their relevance to this thesis.

7.3.1 Generalizability across non-European ancestries

As mentioned throughout this thesis, the most prominent limitation to PRS and genomics in general is the bias towards European ancestry. The vast majority of genetic discovery in GWAS used to train PRSs heavily underrepresent non-European populations - approximately 79% of all GWAS participants are of European descent (237). PRS prediction accuracy decays substantially with increasing genetic divergence from the populations used in the discovery GWAS. For instance, prediction accuracy is, on average, 1.6-fold lower in Hispanic/Latino Americans and South Asians, 2.0-fold lower in East Asians, and 4.5-fold lower in Africans compared to

Europeans. This issue has been widely recognized and efforts are underway to validate existing scores in other ancestries and prioritize of resources for recruitment and data generation in these populations. The GIGASTROKE project for example derived PRS specifically tailored to East Asians, leveraged cross-ancestry GWAS, and attempted to validate models in non-European populations (10). In contrast, the PRS analysis as part of the latest AD-GWAS is comprised entirely of European ancestry (12). Some efforts are being made to expand GWAS in non-European ancestries for dementia (379,380), including cross-ancestry approaches comprising a significant portion of non-Europeans (381,382). However, a recent comprehensive analysis of both European and cross-ancestry AD-PRS did not find substantial improvements for cross-ancestry AD-PRS beyond *APOE* (338). When applying iPRS-DEM to non-European ancestries in *All of Us* we show lack of transferability to Hispanics and African-Americans, and our findings in East Asians, though positive, could be spurious due to low power in this population. A few methodological advances could help close the gap beyond efforts to increase GWAS sample size (383,384). These include improved statistical models that incorporate multi-ancestry summary statistics and reference panels, and to incorporate functional information focused on causal variants that are typically more portable across ancestries (383). In addition, there needs to be more investigation into how social determinants of health may act as effect modifiers or mediators of PRS performance (383).

7.3.2 Lack of standardization in PRS calculation and reporting

Beyond the obvious need to increase GWAS sample sizes and improve the precision of estimated effect sizes, there is also an issue of sample and method dependence resulting in a variability of among PRS for the same disease (238,239). Despite most PRS for the same disease being correlated, there is large discrepancies in individual risk prediction and substantial discordance between different PRS for the same disease at specified risk thresholds, making it uncertain which PRS model is optimal (385). Discrepancies arise from methodological choices regarding algorithms, model complexity, and tuning parameters. Indeed, it has been shown that individuals with an extreme PRS value as calculated by one method may have a PRS value near the middle of the distribution with another (217). A reference-standardized framework can improve the comparability of PRSs across cohorts by ensuring that the calculated PRS is independent of specific properties of the target sample (such as available genetic variation, LD, and MAF) (386). This framework is implemented in *GenoPred* pipeline (386,387), which I used throughout this thesis to standardize PRS calculation in multiple cohorts. Efforts have also been made to standardize PRS reporting to improve reproducibility of findings, and ultimately help determine which PRS is best for a given disease specific to a given population (238). The iPRS-DEM is

accordingly available on the *PGScatalog* so that researchers can validate it in their own datasets in the future.

Once a PRS has demonstrated clinical utility in a given population, laboratories must develop an analytically and clinically valid pipeline for calculating, interpreting and reporting PRS results for an individual patient (320). This step also importantly includes ancestry-based calibration. A comprehensive example of how clinical implementation of PRS might work in practice is provided by the eMERGE study using AoU genetic and EHR data (388). Briefly, optimized PRS across 11 common conditions were put through a technical validation pipeline that included assessing PRS pipeline accuracy where PRS correlations between imputed genotyped data and WGS were compared in the same 70 samples, repeatability where the same pipeline was run ten times over the course of two weeks, and reproducibility measuring three samples, each run six times end-to-end then compared in a pairwise manner (388). Ancestry calibration involved population-based z-score adjustment to account for differences in PRS distribution across ancestries. The thorough efforts of eMERGE consortium should be commended, and provide a framework for standardization of PRS that I believe will go a long way to addressing this limitation.

7.3.3 Limitations in study design: Over-reliance on biobanks, lack of independent cohorts, and phenotype misclassification

As mentioned in Chapter 3 (Methods), a crucial component of valid PRS studies is that the base or discovery GWAS used to derive the SNP-trait effect sizes is entirely independent from the target cohort. This overlap is a critical source of bias because it can result in substantial inflation of the association between the PRS and the tested trait. The magnitude of inflation is proportional to the fraction of the target sample that overlaps the base GWAS, rather than the common, erroneous assumption that inflation is diminished with increasing larger GWAS samples (389,390). From my own experience, we thought at the beginning of constructing iPRS-DEM that there would not be substantial inflation in 3C due to the AD-PRS because 3C only contributed a small fraction of the latest AD-GWAS (12). Yet, association statistics were much larger than expected: HR=2.5-3.0 versus HR=1.1-1.5 found in prior studies. This forced us to systematically re-run GWAS meta-analyses removing 3C and Memento and avoid this inflation. In practice, finding independent cohorts is difficult to achieve. The massive efforts to increase GWAS sample size has led to inclusion of biobanks (especially UKBB), and most large population-based cohorts with the trait of interest. Novel GWAS are also encouraged to include high-risk populations, like memory clinic cohorts (e.g., Memento is included in the latest AD-GWAS). While this does increase sample size and lead to more significant GWAS hits, it makes finding independent validation cohorts for risk

prediction difficult. Recognizing this issue, methods have been developed to address sample overlap by essentially de-correlating summary statistics from the target cohort from the base GWAS, though these methods rely on strict assumptions and recommend the target sample is of sufficient size (N>5,000) (389,390). Beyond sample overlap considerations, there are other important implications in the study design of GWAS, where bigger is almost always considered better.

Beyond the obvious selection towards European-ancestry, there are more subtle selection biases and misspecifications in GWAS that may hinder generalizability and risk prediction (391). GWAS rely heavily on large biobanks that have known participation biases, resulting from the non-random sampling inherent to volunteer-based biobanks. Indeed, UKBB participants tend to be healthier, better educated, and have higher socio-economic status than the general population (392). Study participation itself has a genetic component, and a strong genetic correlation with socio-behavioural traits like educational attainment and income (393). In fact, adjustment for participation bias in the UK Biobank yielded substantial changes to effect sizes across an array of traits, including vascular risk factors like smoking, and BMI (393), but less pronounced changes for molecular traits (393). Subsequently, this bias impacts downstream analyses like heritability estimation, PRS, and MR. Some authors recommend adjustment using inverse probability weighting to correct for this bias, indicating this adjustment significantly reduces the distortion—by approximately 87% on average for binary variable associations and 78% for discrete or continuous variables—thereby making the estimates closer to those found in the representative population (392). However, this correction requires individual-level data and would not be feasible after GWAS summary statistics are published, or across large-scale meta-analysis. It is likely that this bias is pervasive in the summary statistics used in this thesis, especially for traits that are socio-behavioural, and could affect associations with dementia for which educational attainment is an important predictor. It is hard to determine exactly how results may be affected, but one can consider that this bias hindered overall predictive ability and may have slightly distorted associations in population-based cohorts, but even more so in high-risk populations. Participation bias should be more carefully considered at the study design stage for GWAS meta-analyses, and sensitivity analyses should be performed if possible.

In this vein, Escott-Price et al. have strong critiques on the study design of AD-GWAS given the observed reduction in SNP-based heritability: decreasing from approximately 40% in smaller, precise studies to 2–3% in the latest largest GWAS (206). They argue that the current approach to large-scale meta-analyses in AD results in "Russian-doll-like GWAS" where systematic errors

in phenotyping and other pervasive biases are perpetuated and carried forward as sample size increases (206,394). They also levy strong criticism mostly towards diagnostic imprecision UK Biobank and the use of proxy-cases for AD—where self-reported parental history of dementia are counted as cases (206). They estimate that approximately 50% of parental proxy cases may have a different form of dementia (other than AD) or no dementia at all (206). They also highlight inappropriate control selection and age mismatching: only a small fraction of controls are clinically screened (10%), and there is a systematic use of controls that are younger than cases. It is strongly suggested that the genetic architecture of AD depends on the age at clinical onset. Specifically, *APOE* significantly affects longevity and mortality, and adjusting for age as a covariate is insufficient, and can lead to a "robust but biologically spurious association" between AD and age-associated variants (206,395). The consequence of phenotype misclassification and imperfect study design may lead to misleading genetic inferences such as overlapping genetic architecture between AD and Parkinson's Disease likely driven by inclusion of Lewy body dementia cases in AD-GWAS (206). Additionally, these biases likely influence inconsistencies among in individual risk prediction PRS for AD as highlighted above (395). The authors advocate for age-matching of controls, more GWAS studies based specifically on pathologically confirmed AD, endo-phenotypes and biomarkers, and the need to perform age-stratified GWAS (206). Further, sensitivity analyses could be performed to detect the level of phenotype misclassification in GWAS and new phenotyping algorithms could improve effect estimation (396,397)

These are important considerations for the work in this thesis and are meaningful for genetic risk prediction efforts across many common diseases. The main takeaway is that in addition to attempting to improve PRS methodologies with advanced statistical methods, one should first consider the quality of the underlying GWAS and hidden biases that could be addressed to improve genetic risk prediction, especially in diverse sub-groups. Future GWAS should thoughtfully balance between more precise phenotyping and careful study design and the need for larger samples and greater statistical power.

7.3.4 Criticism of the polygenic additive model

Despite advances in PRS and evidence that they provide clinical utility for common diseases when integrated into clinical risk models, they largely fail to capture a substantial portion of heritability. Indeed, PRS for stroke, dementia, and cSVD captured at most 2-3% of the variation across analyses in this thesis. Important criticisms of the polygenic-additive-liability (PAL) model that underlies PRS shed light on the limits of PRS (398,399). The PAL model assumes that binary disease phenotypes are determined by an unobserved quantitative trait, called liability, and that

disease occurs when an individual's liability value exceeds a certain threshold. This model makes several assumptions that are overly simplistic:

i) **Homogeneous genetic determinism:** Diseases have a uniform or single genetic mechanism that explains all forms of the disease. Indeed, as we explain in Chapter 1 and throughout the thesis, cerebrovascular disease (and many other complex diseases) are highly heterogeneous with different forms and pathways, and a single genetic model cannot correctly explain them all.

ii) **Negligible Non-Additive Effects:** Dominance and epistatic effects are negligible.
   a. **Dominance:** risk conferred by being heterozygous for an allele is different from the risk conferred by being homozygous for that allele.
   b. **Epistasis:** Refers to gene-gene interactions.

iii) **Small, independent genetic and environmental effects:** The trait's liability is assumed to depend on the contribution of many weak and independent genetic and environmental factors. Furthermore, the environmental factors are assumed to be normally distributed. This is often criticized because environments change, their effects can be large, and genetic and environmental variance are highly interconnected.

These assumptions can lead to PRSs containing information on both genetic and environmental factors with failure to account for critical non-additive effects, such as dominance and epistasis, all of which are necessary for assessing the absolute risk of developing a disease (398). Consequently, the resulting PRSs likely do not accurately measure the true polygenic predisposition, and may further reflect factors associated with intermediate traits (398). In essence, PRS may not measure what they are supposed to measure (400). Critics state that disease risk cannot be accurately predicted with this over-simplistic model, and further argue that misinterpreting PRS or use of PRS in clinical contexts may lead to unjustified preventive treatment (398). Or worse, they suggest eugenic consequences by accrediting the idea that diseases are genetically determined and one’s genetic risk is known since birth (398).

These criticisms are strong and, in some respects, quite convincing. They provide concrete reasons why PRS have largely not lived up to expectations (401). Yet, empirical data from twin studies and quantitative genetics theory show that most genetic variance in a population is additive (402,403). Largely it is accepted that the PAL model remains useful as the foundation of GWAS to quantify and partition genetic variation in the population and to generate PRS (403). Rather, a more measured approach is to recognize what PRS is actually measuring and to address limitations while avoiding misinterpretations or unrealistic expectations (401). Importantly, these criticisms

should make one reflect that PRS is likely not solely comprised of independent genetic effects, and that the predictive ability is likely intertwined with, or mediated by non-genetic factors (400). For instance, it is encouraged to consider that SNPs are likely associated through intermediate clinical risk factors, which is in fact motivation to perform integrative or pathway-specific PRS that could capture these effects (400). In general, the content validity of PRS should be questioned, in other words, whether the score captures what it ought to measure in a specific clinical, lifestyle, or environmental context.

7.3.5 Barriers to clinical implementation

Considering the methodological and scientific limitations described in sections 7.3.1–7.3.4, a timely, comprehensive review summarizes the barriers to PRS clinical implementation. Major barriers revolve around the lack of evidence base, potential misinterpretation of risk by both providers and patients, and an uncertain regulatory landscape (404).

The most fundamental barrier is the limited empirical evidence that PRS disclosure actually improves health outcomes. Despite growing interest, only a small number of RCTs have evaluated the clinical utility of PRS, and none to date are specific to stroke or dementia. The MI-GENES trial remains the most cited example, demonstrating that disclosure of a coronary artery disease PRS led to modestly lower LDL-cholesterol levels and, in a 10-year follow-up, fewer major cardiovascular events — an effect attributed largely to earlier initiation of statin therapy (405). No equivalent evidence exists for cerebrovascular disease, where clinical actionability following a high-risk PRS result is less clearly defined. Stroke and dementia were notably absent from the eMERGE Network's selection of conditions for clinical PRS implementation, precisely because of this gap in demonstrated clinical validity and actionability. The evidence base for routine clinical use of PRS remains insufficient.

Risk communication represent an equally important challenge. Most clinicians lack formal training in genomics and are unfamiliar with the probabilistic nature of polygenic risk, which differs fundamentally from the binary logic of monogenic test results (406). Communicating a PRS result — including its uncertainty, its dependence on ancestry, and its modifiability through lifestyle and treatment — requires skills that are not yet embedded in routine clinical training (404). Patients, on the other hand, may often hold deterministic views of genetic risk, underestimating the role of modifiable factors, and have cultural beliefs around heredity and fate further that may further complicate risk communication (404). Section 7.4.2 discusses these communication and psychosocial considerations in more depth.

Finally, the regulatory and policy landscape for PRS as clinical tests remains unsettled. A central unresolved question is whether a PRS should be classified as a clinical decision support tool — subject to software validation standards — or as a medical device requiring more stringent regulatory review (404,407). This distinction has meaningful implications for how quickly PRS can be updated as methodology improves and as new GWAS data become available. Broader adoption will also require consensus from professional societies, standardised reporting frameworks, and clear guidelines governing when and how PRS should be offered, interpreted, and acted upon (404). These are not insurmountable challenges, and ongoing initiatives — including the ClinGen PRS Clinical Utility Working Group and international genomics programmes — are actively working toward this goal (404). However, for cerebrovascular disease in particular, where the scientific foundations are still being established, clinical implementation remains a longer-term prospect.

7.4 Perspectives and Future Directions

*7.4.1 **Future Directions:** Epistatic effects, lifestyle and environment*

This thesis started with the focus on improving and exploring PRS models for cerebrovascular diseases. We accomplished the major aims, and provided valuable evidence for the application of PRS for dementia, stroke, and cSVD across different population-based and clinical contexts. However, the overall prediction performance was modest, and PRS for cerebrovascular disease are not ready for clinical use, outside of some potentially very specific high-risk contexts. The limitations outlined in the previous section highlight the technical short-comings and study design biases that impact PRS accuracy and validity. Future research applying PRS, and GWAS more generally, should focus on addressing ancestry bias, and more carefully consider GWAS study design. A stronger focus should be made on precise phenotyping, and an added emphasis on the value of GWAS for biomarkers and endo-phenotypes, like cSVD. The ancestry bias is optimistically being addressed through strong collaborative efforts across the globe, combined with new multi-ancestry methods to maximize signal in diverse ancestries. More recent GWAS are focused on novel biomarkers such as CSF, blood-based, and neuroimaging biomarkers of both AD and cSVD. Essentially, PRS will become more precise and predictive as the quality of the underlying GWAS improves.

We can also investigate more complex PRS that integrate non-additive genetic effects. GenoBoost is one novel method used to integrate non-additive genetic dominance effects and shows particular improvement for prediction of autoimmune diseases (408). Another approach used neural networks to model epistatic effects and GxE interactions in PRS, but found that traditional linear regression models consistently outperformed these models in predicting real traits from the UKBB (409). The authors suggest this could be due to challenges in modeling LD, where technical artifacts that satisfy the mathematical definition of statistical epistasis but do not indicate deeper biological processes, or that there are indeed low levels of non-additive effects (409). Approaches such as these are important attempts to understand the contribution of non-additive effects and their potential for genetic risk prediction.

In addition, through this work, we found that considering PRS in the context of other important genetic factors (e.g. *APOE* in dementia), or deconstructing risk into biological pathways provided important insights. Further integration of additional multi-omics data such as proteomics, transcriptomics, and metabolomics is a natural next step as well. Recently, using data from the INTERVAL study, an atlas of PRS for over 17,000 multi-omic traits including proteins, metabolites, and RNA transcripts was generated (410). The use of these PRS as proxies for multi-

omics traits bypasses the time and costs for collecting multi-omics data, and presents a starting point to identify multi-omic predictors of disease and implicated pathways (410). In another study, both gene expression (i.e., transcriptomics) and methylation data are combined in interpretable neural networks, and showed strong predictive ability for smoking status (411). For cSVD, in particular, investigations leveraging PRS and other multi-omics layers together, where available, may provide valuable insights into disease mechanisms that may eventually inform prediction models.

It will eventually be necessary to incorporate PRS with established clinical, environmental, and lifestyle factors in integrated risk tools. A promising avenue of future research is investigating PRS in the context of the exposome—or the totality of all environmental exposures. These explorations have potential to advance our understanding of how genetics interacts with the environment which will have an ultimately positive impact on risk prediction models. One study attempted to parse direct and indirect genetic effects that may influence PRS portability across sub-groups beyond ancestry differences (412). In particular, this study highlights varying PRS prediction performance in sub-groups according to sex, and socio-economic status. Further explorations between standard and family-based GWAS attribute this in part to indirect genetic effects of parents, siblings or peers, or assortative mating—a process where individuals are more likely to mate if they share the same phenotype (412). This study ultimately underscores that PRS accuracy is sensitive to the environmental and social context in which the GWAS was performed. Another important study in the UK Biobank determined that, beyond age and sex, the exposome accounts for a substantially larger proportion of variation in biological aging measured by a proteomic clock and all-cause mortality (~17%) compared to genetic factors represented by a sets of PRS (~2%) (413). However, for some age-related conditions, notably dementia, PRS captured a larger proportion of variance, whereas for heart disease environmental factors were stronger (413). This study did not investigate potential PRS-environment interaction, and note this as an important future direction.

In this vein, advances have been made to directly model GxE interactions in a PRS model. A recent innovative study employed such models and detected significant GxE interactions in quantitative obesity-related traits, and also disease-specific GxE interactions (414). For example the genetic effects on stroke were influenced by waist-hip ratio (414). Importantly, they also provide a multi-phase framework with different configurations to model environmental variables in GxE-PRS models adaptable to different research questions and contexts for genetic risk prediction. Another novel approach investigated pathway-based PRS by environment interactions and showed this

approach significantly increased statistical power to detect interaction effects compared to a global PRS (415). By characterizing pathways with functional relevance with ps-PRS, this approach can identify mechanisms of action relating an environmental exposure, or certain drugs, to a disease. In fact, this represents an interesting extension to the work presented in Article 3.

Investigating PRS in the context of environmental factors by exploiting these novel approaches may go a long way to informing preventive strategies especially for multi-modal interventions and promotion of healthier life styles. It remains that multimodal interventions targeting behavioural, lifestyle and environmental factors along with management of vascular risk factors are the most effective form of prevention for cerebrovascular diseases. Incorporating genetic risk prediction to support primary prevention strategies, including lifestyle modification, is crucial, as is the continued identification of therapeutic targets and biological mechanisms through multi-omics investigations.

*7.4.2 **Perspective:** Effective risk communication is essential to PRS utility*

The clinical utility of PRS also rely on effective risk communication between providers and patients, and the complementary use of genetic counseling. In addition, the personal utility of PRS has been highlighted to empower individuals in their health management and decision-making, and could motivate positive lifestyle modifications and inform life planning (18). These potential benefits need to be balanced against negative psychosocial impacts like anxiety, fatalistic thinking or a false sense of security (18). At the crux of this effective risk communication is understanding the inherent uncertainty in the risk estimated by PRS, and health care providers will require additional education to this end (240). Ultimately, the goal of this communication should be to motivate patients toward risk-reducing health behaviors and minimize potential negative psychosocial effects such as anxiety or depression (240,416).

A 2016 meta-analysis across 18 trials found no significant effect of genetic risk information on motivating changes in behaviors like quitting smoking, improving diet, or increasing physical activity (417). Another more recent study found that disclosure of CAD risk, including polygenic and clinical factors, did not result in meaningful changes to lifestyle or biological risk factors (418). From a psychosocial perspective, a small RCT investigating the impact of returning T2D and CAD genetic risk estimates to patients, did not find significant psychosocial harm beyond returning risk based on traditional risk factors (419). There is also limited evidence that suggests there is a potential benefit in providing people with PRS results in CAD prevention (325,420), which could increase participants' perception of personal control, information seeking, and greater initiation of statins as a result of shared decision making between patients and providers (325,405).

However, more research is required to determine how PRS results influence clinicians' recommendations and patients' decision making, and how they impact risk-reducing health behaviors, especially among younger adults (325). Presently, eMERGE (electronic MEdical Records and GEnomics) is a large-scale prospective cohort study enrolling 25,000 individuals designed to investigate the utility of returning a comprehensive report based on electronic medical records, integrating clinical and genetic risk across 11 complex conditions – including vascular risk factors and different cancers (421). Specifically, they aim to determine if returning this information is effective at increasing appropriate clinical actions and uptake of preventive therapies to mitigate future disease risk (421). Notably, both stroke and dementia were not among selected complex disease, likely due to the lack of demonstrated clinical validity of PRS for these conditions, and limited clinical actionability as discussed above. Still, the results of these studies will go a long way to informing the integration of genetic risk in clinical practice.

Direct-to-Consumer Genetic Testing (DTC-GT), including monogenic mutations and PRS, is rapidly expanding and beginning to be widely accessible in North America despite the general lack of sufficient evidence to support its use in routine clinical care (422). In recent years, the Food and Drug Administration (FDA) has approved marketing of DTC-GTs for late-onset AD among some other diseases. In light of this, the *American Heart Association* released a scientific statement regarding DTC-GT in cardiovascular disease (422). They summarize that although "*the expansive market may democratize access to genetic information, the insufficient evidence for its use in routine clinical care, combined with limited oversight and variability in testing methods, generates confusion for both patients and clinicians*" (422). While DTC-GT offers benefits like increased consumer autonomy and accessibility, these are counterbalanced by risks like misinterpretation, test inaccuracy, and the lack of standardized systems for reporting results and subsequent genetic counseling (422). It's clear that with the expanding market of DTC-GTs that there will be pressure on the medical, legal, and research community to establish guidelines for use of PRS across common diseases, and to improve PRS and address known limitations as fast as possible.

7.5 General conclusion

Preventing cerebrovascular disease in the coming decades will be essential for safeguarding the health of aging populations and maintaining sustainable healthcare systems. PRS have the potential to contribute meaningfully to this effort by enabling earlier identification of high-risk individuals, informing clinical decision-making, and enriching our understanding of disease processes. In this thesis, I examined cerebrovascular disease—from cSVD to stroke and dementia—across diverse contexts, including the general population, multiple-ancestry cohorts, and high-risk clinical groups such as memory-clinic patients and ICH survivors with AF. We highlight the potential value of integrative PRS that captures vascular and neurodegenerative genetic risk to refine dementia risk stratification, and showed that PRS may help guide anticoagulation strategies in secondary stroke prevention. We further explored the etiology of cSVD using pathway-specific PRS, providing evidence for multiple biological pathways operating across the lifespan and contributing to stroke and dementia in later life. While PRS for cerebrovascular diseases appear most immediately relevant in high-risk clinical settings—where clearer guidelines and more favorable risk–benefit trade-offs exist—their current predictive performance remains insufficient for broader primary prevention. The most effective primary prevention remains multimodal interventions and control of vascular risk factors like hypertension, and PRS has yet to be evaluated in these contexts. Significant challenges remain, including limitations in GWAS quality and inherent biases, particularly the underrepresentation of non-European populations, along with concerns regarding phenotype misclassification and participation bias. PRS also face technical barriers like limited replicability across PRS for the same disease, and insufficient modeling of non-additive genetic or gene–environment interactions. Addressing these shortcomings is essential to ensure that PRS predict well enough for future clinical applications and do not exacerbate existing health disparities. Greater consensus and established guidelines will be needed to standardize PRS construction, reporting, clinical implementation, and risk communication. Nonetheless, multiple promising avenues lie ahead, including multi-omics integration, evaluation of PRS across diverse environmental contexts, and the development of comprehensive prediction frameworks that combine genetics with novel biomarkers and factors capturing the broader human exposome. Although PRS remain controversial, the trajectory of ongoing research suggests that genetic risk prediction may yet hold substantial value for improving the prevention of stroke, dementia, and other common diseases.

References

1. Feigin VL, Abate MD, Abate YH, ElHafeez SA, Abd-Allah F, Abdelalim A, et al. Global, regional, and national burden of stroke and its risk factors, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. [cited 2025 Aug 7]; Available from: https://www.thelancet.com/journals/laneur/article/PIIS1474-4422(24)00369-7/fulltext
2. Nichols E, Vos T. The estimation of the global prevalence of dementia from 1990-2019 and forecasted prevalence through 2050: An analysis for the Global Burden of Disease (GBD) study 2019. Alzheimer's & Dementia. 2021;17(S10):e051496.
3. Cheng Y, Lin Y, Shi H, Cheng M, Zhang B, Liu X, et al. Projections of the Stroke Burden at the Global, Regional, and National Levels up to 2050 Based on the Global Burden of Disease Study 2021. Journal of the American Heart Association. 2024 Dec 3;13(23):e036142.
4. Snyder HM, Corriveau RA, Craft S, Faber JE, Greenberg SM, Knopman D, et al. Vascular contributions to cognitive impairment and dementia including Alzheimer's disease. Alzheimer's & Dementia. 2015 Jun;11(6):710–7.
5. Schneider JA, Arvanitakis Z, Bang W, Bennett DA. Mixed brain pathologies account for most dementia cases in community-dwelling older persons. Neurology. 2007 Dec 11;69(24):2197–204.
6. Gladman JT, Corriveau RA, Debette S, Dichgans M, Greenberg SM, Sachdev PS, et al. Vascular contributions to cognitive impairment and dementia: Research consortia that focus on etiology and treatable targets to lessen the burden of dementia worldwide. Alzheimer's & Dementia: Translational Research & Clinical Interventions. 2019 Jan 1;5:789–96.
7. White matter hyperintensities in vascular contributions to cognitive impairment and dementia (VCID): Knowledge gaps and opportunities. [cited 2025 Feb 12]; Available from: https://www.sciencedirect.com/science/article/pii/S2352873719300046
8. Debette S, Markus HS. The clinical importance of white matter hyperintensities on brain magnetic resonance imaging: systematic review and meta-analysis. BMJ [Internet]. 2010 Jul 26 [cited 2025 Feb 12]; Available from: https://www.bmj.com/content/341/bmj.c3666.abstract
9. Wardlaw JM, Debette S, Jokinen H, De Leeuw FE, Pantoni L, Chabriat H, et al. ESO Guideline on covert cerebral small vessel disease. European Stroke Journal. 2021 Jun 1;6(2):CXI–CLXII.
10. Mishra A, Malik R, Hachiya T, Jürgenson T, Namba S, Posner DC, et al. Stroke genetics informs drug discovery and risk prediction across ancestries. Nature. 2022 Nov;611(7934):115–23.
11. Woo D, Falcone GJ, Devan WJ, Brown WM, Biffi A, Howard TD, et al. Meta-analysis of Genome-wide Association Studies Identifies 1q22 as a Susceptibility Locus for Intracerebral Hemorrhage. The American Journal of Human Genetics. 2014 Apr 3;94(4):511–21.
12. Bellenguez C, Küçükali F, Jansen IE, Kleineidam L, Moreno-Grau S, Amin N, et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. Nature Genetics. 2022 Apr 1;54(4):412–36.
13. Sargurupremraj M, Suzuki H, Jian X, Sarnowski C, Evans TE, Bis JC, et al. Cerebral small vessel disease genomics and its implications across the lifespan. Nat Commun. 2020 Dec 8;11(1):6285.
14. Duperron MG, Knol MJ, Le Grand Q, Evans TE, Mishra A, Tsuchida A, et al. Genomics of perivascular space burden unravels early mechanisms of cerebral small vessel disease. Nat Med. 2023 Apr;29(4):950–62.
15. Knol MJ, Lu D, Traylor M, Adams HHH, Romero JRJ, Smith AV, et al. Association of common genetic variants with brain microbleeds. Neurology. 2020 Dec 15;95(24):e3331–43.
16. Bordes C, Sargurupremraj M, Mishra A, Debette S. Genetics of common cerebral small vessel disease. Nature Reviews Neurology. 2022 Feb;18(2):84–101.
17. Lambert SA, Abraham G, Inouye M. Towards clinical utility of polygenic risk scores. Human Molecular Genetics. 2019 Nov 21;28(R2):R133–42.
18. Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. Nat Rev Genet. 2018 Sep;19(9):581–90.
19. Harrison JR, Mistry S, Muskett N, Escott-Price V. From Polygenic Scores to Precision Medicine in Alzheimer's Disease: A Systematic Review. Journal of Alzheimer's Disease. 2020 Jan 1;74(4):1271–83.
20. Clark K, Leung YY, Lee WP, Voight B, Wang LS. Polygenic Risk Scores in Alzheimer's Disease Genetics: Methodology, Applications, Inclusion, and Diversity. JAD. 2022 Aug 30;89(1):1–12.

21. Abraham G, Rutten-Jacobs L, Inouye M. Risk Prediction Using Polygenic Risk Scores for Prevention of Stroke and Other Cardiovascular Diseases. Stroke. 2021 Sep;52(9):2983–91.

22. O'Sullivan JW, Ashley EA, Elliott PM. Polygenic risk scores for the prediction of cardiometabolic disease. Eur Heart J. 2023 Jan 7;44(2):89–99.

23. Nichols E, Steinmetz JD, Vollset SE, Fukutaki K, Chalek J, Abd-Allah F, et al. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. The Lancet Public Health. 2022 Feb;7(2):e105–25.

24. Vollset SE, Ababneh HS, Abate YH, Abbafati C, Abbasgholizadeh R, Abbasian M, et al. Burden of disease scenarios for 204 countries and territories, 2022–2050: a forecasting analysis for the Global Burden of Disease Study 2021. [cited 2025 Aug 27]; Available from: https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(24)00685-8/fulltext

25. Seshadri S, Beiser A, Kelly-Hayes M, Kase CS, Au R, Kannel WB, et al. The Lifetime Risk of Stroke. Stroke. 2006 Feb;37(2):345–50.

26. Aging and Health fact sheet [Internet]. World Health Organization; 2024. Available from: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health

27. Aging: Global Population, Q&A [Internet]. World Health Organization; 2025. Available from: https://www.who.int/news-room/questions-and-answers/item/population-ageing

28. Olshansky SJ. From Lifespan to Healthspan. JAMA. 2018 Oct 2;320(13):1323–4.

29. Fries JF. The Compression of Morbidity. Milbank Q. 2005 Dec;83(4):801–23.

30. Crimmins EM. Lifespan and Healthspan: Past, Present, and Promise. Gerontologist. 2015 Dec;55(6):901–11.

31. Garmany A, Terzic A. Global Healthspan-Lifespan Gaps Among 183 World Health Organization Member States. JAMA Netw Open. 2024 Dec 11;7(12):e2450241.

32. Richardson C, Stephan BCM, Robinson L, Brayne C, Matthews FE. Two-decade change in prevalence of cognitive impairment in the UK. European Journal of Epidemiology. 2019 Sep 5;34(11):1085–92.

33. Danaei G, Finucane MM, Lin JK, Singh GM, Paciorek CJ, Cowan MJ, et al. National, regional, and global trends in systolic blood pressure since 1980: systematic analysis of health examination surveys and epidemiological studies with 786 country-years and 5·4 million participants. [cited 2025 Aug 27]; Available from: https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(10)62036-3/abstract

34. Ong KL, Stafford LK, McLaughlin SA, Boyko EJ, Vollset SE, Smith AE, et al. Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021. [cited 2025 Aug 28]; Available from: https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(23)01301-6/fulltext

35. Padwal RS. Obesity, Diabetes, and the Metabolic Syndrome: The Global Scourge. [cited 2025 Aug 28]; Available from: https://onlinecjc.ca/article/S0828-282X(13)01635-8/abstract

36. Schutte AE, Srinivasapura Venkateshmurthy N, Mohan S, Prabhakaran D. Hypertension in Low- and Middle-Income Countries. Circulation Research. 2021 Apr 2;128(7):808–26.

37. Crimmins E, Kim JK, Vasunilashorn S. Biodemography: New approaches to understanding trends and differences in population health and mortality. Demography. 2010 Mar 1;47(Suppl 1):S41–64.

38. Herculano-Houzel S. The Human Brain in Numbers: A Linearly Scaled-up Primate Brain. Front Hum Neurosci. 2009 Nov 9;3:31.

39. Wen Q, Chklovskii DB. Segregation of the Brain into Gray and White Matter: A Design Minimizing Conduction Delays. [cited 2025 Oct 13]; Available from: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.0010078

40. Buyanova IS, Arsalidou M. Cerebral White Matter Myelination and Relations to Age, Gender, and Cognition: A Selective Review. Front Hum Neurosci. 2021 Jul 6;15:662031.

41. Filley CM, Fields RD. White matter and cognition: making the connection. J Neurophysiol. 2016 Nov 1;116(5):2093–104.

42. Bear MF, Connors BW, Paradiso MA. Neuroscience: Exploring the brain, 3rd ed. Philadelphia, PA, US: Lippincott Williams & Wilkins Publishers; 2007. xxxviii, 857 p. (Neuroscience: Exploring the brain, 3rd ed).

43. Mannarelli D, Pauletti C, Missori P, Trompetto C, Cotellessa F, Fattapposta F, et al. Cerebellum's Contribution to Attention, Executive Functions and Timing: Psychophysiological Evidence from Event-Related Potentials. 2023;

44. Shahid Z, Asuka E, Singh G. Physiology, Hypothalamus. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 [cited 2025 Oct 14]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK535380/

45. Cb S, Te S, J L. Hypothalamic regulation of sleep and circadian rhythms. PubMed [Internet]. 2005 [cited 2025 Oct 14]; Available from: https://pubmed.ncbi.nlm.nih.gov/16251950/

46. Mergenthaler P, Lindauer U, Dienel GA, Meisel A. Sugar for the brain: the role of glucose in physiological and pathological brain function. Trends Neurosci. 2013 Oct;36(10):587–97.

47. Faria-Pereira A, Morais VA. Synapses: The Brain's Energy-Demanding Sites. Int J Mol Sci. 2022 Mar 26;23(7):3627.

48. Konan LM, Reddy V, Mesfin FB. Neuroanatomy, Cerebral Blood Supply. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 [cited 2025 Oct 14]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK532297/

49. Rosner J, Reddy V, Lui F. Neuroanatomy, Circle of Willis. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 [cited 2025 Aug 31]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK534861/

50. Vrselja Z, Brkic H, Mrdenovic S, Radic R, Curic G. Function of Circle of Willis. J Cereb Blood Flow Metab. 2014 Apr 1;34(4):578–84.

51. Chandra A, Li WA, Stone CR, Geng X, Ding Y. The cerebral circulation and cerebrovascular disease I: Anatomy. Brain Circ. 2017;3(2):45–56.

52. Hufnagle JJ, Tadi P. Neuroanatomy, Brain Veins. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 [cited 2025 Oct 14]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK546605/

53. Iadecola C. The Neurovascular Unit Coming of Age: A Journey through Neurovascular Coupling in Health and Disease. [cited 2025 Aug 29]; Available from: https://www.cell.com/neuron/abstract/S0896-6273(17)30652-9

54. Soto-Rojas LO, Pacheco-Herrero M, Martínez-Gómez PA, Campa-Córdoba BB, Apátiga-Pérez R, Villegas-Rojas MM, et al. The Neurovascular Unit Dysfunction in Alzheimer's Disease. Int J Mol Sci. 2021;

55. Keaney J, Walsh DM, O'Malley T, Hudson N, Crosbie DE, Loftus T, et al. Autoregulated paracellular clearance of amyloid-β across the blood-brain barrier. Science Advances. 2015 Sep 18;1(8):e1500472.

56. Sato Y, Falcone-Juengert J, Tominaga T, Su H, Liu J. Remodeling of the Neurovascular Unit Following Cerebral Ischemia and Hemorrhage. Cells. 2022 Sep 9;11(18):2823.

57. Gottesman RF, Seshadri S. Risk Factors, Lifestyle Behaviors, and Vascular Brain Health. Stroke. 2022 Feb;53(2):394–403.

58. van der Flier WM, Skoog I, Schneider JA, Pantoni L, Mok V, Chen CLH, et al. Vascular cognitive impairment. Nat Rev Dis Primers. 2018 Feb 15;4(1):1–16.

59. Inoue Y, Shue F, Guojun B, Kanekiyo T. Pathophysiology and probable etiology of cerebral small vessel disease in vascular dementia and Alzheimer's disease.

60. Pantoni L. Cerebral small vessel disease: from pathogenesis and clinical characteristics to therapeutic challenges. The Lancet Neurology. 2010 Jul 1;9(7):689–701.

61. Wardlaw JM, Smith C, Dichgans M. Small vessel disease: mechanisms and clinical implications. The Lancet Neurology. 2019 Jul 1;18(7):684–96.

62. Markus HS, Leeuw FE de. Cerebral small vessel disease: Recent advances and future directions. https://doi.org/101177/17474930221144911 [Internet]. 2022 Dec 27 [cited 2025 Feb 12]; Available from: https://journals.sagepub.com/doi/full/10.1177/17474930221144911

63. Debette S, Schilling S, Duperron MG, Larsson SC, Markus HS. Clinical Significance of Magnetic Resonance Imaging Markers of Vascular Brain Injury: A Systematic Review and Meta-analysis. JAMA Neurology. 2019 Jan 1;76(1):81–94.

64. Georgakis MK, Duering M, Wardlaw JM, Dichgans M. WMH and long-term outcomes in ischemic stroke. Neurology. 2019 Mar 19;92(12):e1298–308.

65. Duering M, Biessels GJ, Brodtmann A, Chen C, Cordonnier C, Leeuw FE de, et al. Neuroimaging standards for research into small vessel disease—advances since 2013. [cited 2025 Feb 12]; Available from: https://www.thelancet.com/journals/laneur/article/PIIS1474-4422(23)00131-X/abstract

66. Zhu YC, Dufouil C, Tzourio C, Chabriat H. Silent Brain Infarcts. Stroke. 2011 Apr;42(4):1140–5.

67. Shoamanesh A, Kwok CS, Benavente O. Cerebral Microbleeds: Histopathological Correlation of Neuroimaging. [cited 2025 Oct 20]; Available from: https://dx.doi.org/10.1159/000331466

68. Wardlaw JM, Benveniste H, Nedergaard M, Zlokovic BV, Mestre H, Lee H, et al. Perivascular spaces in the brain: anatomy, physiology and pathology. Nature Reviews Neurology. 2020 Mar 1;16(3):137–53.

69. Wardlaw JM, Smith EE, Biessels GJ, Cordonnier C, Fazekas F, Frayne R, et al. Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration. Lancet Neurol. 2013 Aug;12(8):822–38.

70. Mishra A, Duplaà C, Vojinovic D, Suzuki H, Sargurupremraj M, Zilhão NR, et al. Gene-mapping study of extremes of cerebral small vessel disease reveals TRIM47 as a strong candidate. Brain. 2022 Jun 1;145(6):1992–2007.

71. Le Grand Q, Tsuchida A, Koch A, Imtiaz MA, Aziz NA, Vigneron C, et al. Diffusion imaging genomics provides novel insight into early mechanisms of cerebral small vessel disease. Molecular Psychiatry. 2024 Nov;29(11):3567–79.

72. Baykara E, Gesierich B, Adam R, Tuladhar AM, Biesbroek JM, Koek HL, et al. A Novel Imaging Marker for Small Vessel Disease Based on Skeletonization of White Matter Tracts and Diffusion Histograms. Annals of Neurology. 2016;80(4):581–92.

73. Croall ID, Lohner V, Moynihan B, Khan U, Hassan A, O'Brien JT, et al. Using DTI to assess white matter microstructure in cerebral small vessel disease (SVD) in multicentre studies. Clinical Science. 2017 Jun 7;131(12):1361–73.

74. Zhang H, Schneider T, Wheeler-Kingshott CA, Alexander DC. NODDI: Practical *in vivo* neurite orientation dispersion and density imaging of the human brain. NeuroImage. 2012 Jul 16;61(4):1000–16.

75. Markus HS, Joutel A. The pathogenesis of cerebral small vessel diseases and vascular cognitive impairment. https://doi.org/101152/physrev000282024 [Internet]. 2025 Feb 18 [cited 2025 Sep 19]; Available from: https://journals.physiology.org/doi/10.1152/physrev.00028.2024

76. Wardlaw JM, Allerhand M, Doubal FN, Valdes Hernandez M, Morris Z, Gow AJ, et al. Vascular risk factors, large-artery atheroma, and brain white matter hyperintensities. Neurology. 2014 Apr 15;82(15):1331–8.

77. Backhouse EV, Boardman JP, Wardlaw JM. Cerebral Small Vessel Disease: Early-Life Antecedents and Long-Term Implications for the Brain, Aging, Stroke, and Dementia. Hypertension. 2024 Jan;81(1):54–74.

78. Bath PM, Wardlaw JM. Pharmacological Treatment and Prevention of Cerebral Small Vessel Disease: A Review of Potential Interventions. International Journal of Stroke. 2015 Jun 1;10(4):469–78.

79. Group TSR. The New England Journal of Medicine. Massachusetts Medical Society; 2015 [cited 2025 Oct 20]. A Randomized Trial of Intensive versus Standard Blood-Pressure Control. Available from: https://www.nejm.org/doi/full/10.1056/NEJMoa1511939

80. Group TSMI for the SR, Nasrallah IM, Pajewski NM, Auchus AP, Chelune G, Cheung AK, et al. Association Between Intensive vs Standard Blood Pressure Control and Cerebral White Matter Lesions. [cited 2025 Oct 20]; Available from: https://jamanetwork.com/journals/jama/fullarticle/2747671

81. Group TSMI for the SR, Williamson JD, Pajewski NM, Auchus AP, Bryan RN, Chelune G, et al. Effect of Intensive vs Standard Blood Pressure Control on Probable Dementia. [cited 2025 Oct 20]; Available from: https://jamanetwork.com/journals/jama/fullarticle/2723256

82. K A, P H, S H, J M, Ve S, T S. Cerebrovascular disease in the community: results of a WHO collaborative study. PubMed [Internet]. [cited 2025 Oct 15]; Available from: https://pubmed.ncbi.nlm.nih.gov/6966542/

83. Hilkens NA, Casolla B, Leung TW, Leeuw FE de. Stroke. [cited 2025 Oct 14]; Available from: https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(24)00642-1/abstract

84. Adams HP, Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, et al. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. Stroke. 1993 Jan;24(1):35–41.

85. O'Donnell MJ, Xavier D, Liu L, Zhang H, Chin SL, Rao-Melacini P, et al. Risk factors for ischaemic and intracerebral haemorrhagic stroke in 22 countries (the INTERSTROKE study): a case-control study. [cited 2025 Oct 14]; Available from: https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(10)60834-3/abstract

86. Cordonnier C, Demchuk A, Ziai W, Anderson CS. Intracerebral haemorrhage: current approaches to acute management. [cited 2025 Oct 16]; Available from: https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(18)31878-6/abstract

87. Ekker MS, Verhoeven JI, Vaartjes I, van Nieuwenhuizen KM, Klijn CJM, de Leeuw FE. Stroke incidence in young adults according to age, subtype, sex, and time trends. Neurology. 2019 May 21;92(21):e2444–54.

88. Diener HC, Hankey GJ. Primary and Secondary Prevention of Ischemic Stroke and Cerebral Hemorrhage. JACC. 2020 Apr 21;75(15):1804–18.

89. Owolabi MO, Thrift AG, Mahal A, Ishida M, Martins S, Johnson WD, et al. Primary stroke prevention worldwide: translating evidence into action. [cited 2025 Oct 16]; Available from: https://www.thelancet.com/journals/lanpub/article/PIIS2468-2667(21)00230-9/fulltext

90. Law MR, Morris JK, Wald NJ. Use of blood pressure lowering drugs in the prevention of cardiovascular disease: meta-analysis of 147 randomised trials in the context of expectations from prospective epidemiological studies. 2009 May 19 [cited 2025 Oct 16]; Available from: https://www.bmj.com/content/338/bmj.b1665

91. Ruff CT, Giugliano RP, Braunwald E, Hoffman EB, Deenadayalu N, Ezekowitz MD, et al. Comparison of the efficacy and safety of new oral anticoagulants with warfarin in patients with atrial fibrillation: a meta-analysis of randomised trials. [cited 2025 Oct 16]; Available from: https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(13)62343-0/abstract

92. Mohan KM, Wolfe CDA, Rudd AG, Heuschmann PU, Kolominsky-Rabas PL, Grieve AP. Risk and Cumulative Risk of Stroke Recurrence. Stroke. 2011 May;42(5):1489–94.

93. Kleindorfer DO, Towfighi A, Chaturvedi S, Cockroft KM, Gutierrez J, Lombardi-Hill D, et al. 2021 Guideline for the Prevention of Stroke in Patients With Stroke and Transient Ischemic Attack: A Guideline From the American Heart Association/American Stroke Association. Stroke. 2021 Jul;52(7):e364–467.

94. Feigin VL, Owolabi MO, Feigin VL, Abd-Allah F, Akinyemi RO, Bhattacharjee NV, et al. Pragmatic solutions to reduce the global burden of stroke: a World Stroke Organization–Lancet Neurology Commission. [cited 2025 Oct 14]; Available from: https://www.thelancet.com/journals/laneur/article/PIIS1474-4422(23)00277-6/fulltext

95. Zhang R, Wang Y, Fang J, Yu M, Wang Y, Liu G. Worldwide 1-month case fatality of ischaemic stroke and the temporal trend. Stroke and Vascular Neurology [Internet]. 2020 Dec 29 [cited 2025 Oct 16]; Available from: https://svn.bmj.com/content/5/4/353

96. Wolsink A, Cliteur MP, Asch CJ van, Boogaarts HD, Dammers R, Hannink G, et al. Incidence, case fatality, and functional outcome of intracerebral haemorrhage, according to age, sex, and country income level: a systematic review and meta-analysis. [cited 2025 Oct 16]; Available from: https://www.thelancet.com/journals/lanepe/article/PIIS2666-7762(24)00349-1/fulltext

97. Romain G, Mariet AS, Jooste V, Duloquin G, Thomas Q, Durier J, et al. Long-Term Relative Survival after Stroke: The Dijon Stroke Registry. Neuroepidemiology. 2019 Dec 19;54(6):498–505.

98. Greenberg SM, Ziai WC, Cordonnier C, Dowlatshahi D, Francis B, Goldstein JN, et al. 2022 Guideline for the Management of Patients With Spontaneous Intracerebral Hemorrhage: A Guideline From the American Heart Association/American Stroke Association. Stroke. 2022 Jul;53(7):e282–361.

99. Li L, Poon MTC, Samarasekera NE, Perry LA, Moullaali TJ, Rodrigues MA, et al. Risks of recurrent stroke and all serious vascular events after spontaneous intracerebral haemorrhage: pooled analyses of two population-based studies. [cited 2025 Oct 16]; Available from: https://www.thelancet.com/journals/laneur/article/PIIS1474-4422(21)00075-2/fulltext

100. Sheth KN, Selim M. Focused Update on Vascular Risk and Secondary Prevention in Survivors of Intracerebral Hemorrhage. Stroke. 2022 Jul;53(7):2128–30.

101. Salman RAS, Dennis MS, Sandercock P a. G, Sudlow CLM, Wardlaw JM, Whiteley WN, et al. Effects of antiplatelet therapy after stroke due to intracerebral haemorrhage (RESTART): a randomised, open-label trial. [cited 2025 Oct 16]; Available from: https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(19)30840-2/fulltext

102. Veltkamp R, Korompoki E, Harvey KH, Harvey ER, Fießler C, Malzahn U, et al. Direct oral anticoagulants versus no anticoagulation for the prevention of stroke in survivors of intracerebral haemorrhage with atrial fibrillation (PRESTIGE-AF): a multicentre, open-label, randomised, phase 3 trial. The Lancet. 2025 Mar 15;405(10482):927–36.

103. Kaisaier W, Guan C, Ye M, Guo S, Chen Y, Dong Y, et al. Effect of oral anticoagulant therapy on adverse outcomes in patients with atrial fibrillation after intracranial hemorrhage. [cited 2025 Sep 1]; Available from: https://www.heartrhythmjournal.com/article/S1547-5271(25)02602-5/fulltext

104. Toffali M, Carbone F, Fainardi E, Morotti A, Montecucco F, Liberale L, et al. Secondary prevention after intracerebral haemorrhage. European Journal of Clinical Investigation. 2023;53(6):e13962.

105. Petersen RC. Mild cognitive impairment as a diagnostic entity. Journal of Internal Medicine. 2004;256(3):183–94.

106. Emmady PD, Schoo C, Tadi P. Major Neurocognitive Disorder (Dementia). In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 [cited 2025 Oct 18]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK557444/

107. Diagnostic and statistical manual of mental disorders. 5th ed. American Psychiatric Association; 2013.

108. Hugo J, Ganguli M. Dementia and Cognitive Impairment. [cited 2025 Oct 18]; Available from: https://www.geriatric.theclinics.com/article/S0749-0690(14)00036-6/abstract

109. Elahi FM, Miller BL. A clinicopathological approach to the diagnosis of dementia. Nature Reviews Neurology. 2017 Aug;13(8):457–76.

110. Gorelick PB, Scuteri A, Black SE, DeCarli C, Greenberg SM, Iadecola C, et al. Vascular Contributions to Cognitive Impairment and Dementia. Stroke. 2011 Sep;42(9):2672–713.

111. Hendriks S, Peetoom K, Bakker C, van der Flier WM, Papma JM, Koopmans R, et al. Global Prevalence of Young-Onset Dementia: A Systematic Review and Meta-analysis. JAMA Neurol. 2021 Sep 1;78(9):1080–90.

112. Walker L, McAleese KE, Thomas AJ, Johnson M, Martin-Ruiz C, Parker C, et al. Neuropathologically mixed Alzheimer's and Lewy body disease: burden of pathological protein aggregates differs between clinical phenotypes. Acta Neuropathologica. 2015 Mar 11;129(5):729–48.

113. Kehagia AA, Barker RA, Robbins TW. Neuropsychological and clinical heterogeneity of cognitive impairment and dementia in patients with Parkinson's disease. [cited 2025 Oct 19]; Available from: https://www.thelancet.com/journals/laneur/article/PIIS1474-4422(10)70212-X/abstract

114. Alzheimer's Research UK. Subtypes of dementia [Internet]. Dementia Statistics Hub; [cited 2025 Oct 19]. Available from: https://dementiastatistics.org/about-dementia/subtypes/

115. Knopman DS, Amieva H, Petersen RC, Chételat G, Holtzman DM, Hyman BT, et al. Alzheimer disease. Nature Reviews Disease Primers. 2021 May 13;7(1):33.

116. Jack Jr. CR, Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, et al. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. Alzheimer's & Dementia. 2018;14(4):535–62.

117. Makin S. The amyloid hypothesis on trial. Nature. 2018 Jul 25;559(7715):S4–7.

118. Jack CR Jr, Therneau TM, Weigand SD, Wiste HJ, Knopman DS, Vemuri P, et al. Prevalence of Biologically vs Clinically Defined Alzheimer Spectrum Entities Using the National Institute on Aging–Alzheimer's Association Research Framework. JAMA Neurol. 2019 Oct 1;76(10):1174–83.

119. Frisoni GB, Altomare D, Thal DR, Ribaldi F, van der Kant R, Ossenkoppele R, et al. The probabilistic model of Alzheimer disease: the amyloid hypothesis revised. Nature Reviews Neuroscience. 2022 Jan;23(1):53–66.

120. Kepp KP, Robakis NK, Høilund-Carlsen PF, Sensi SL, Vissel B. The amyloid cascade hypothesis: an updated critical review. Brain. 2023 Oct 3;146(10):3969–90.

121. Pathological correlates of late-onset dementia in a multicentre, community-based population in England and Wales. The Lancet. 2001 Jan 20;357(9251):169–75.

122. Mok VCT, Cai Y, Markus HS. Vascular cognitive impairment and dementia: Mechanisms, treatment, and future directions. International Journal of Stroke. 2024 Oct 1;19(8):838–56.

123. Sexton E, McLoughlin A, Williams DJ, Merriman NA, Donnelly N, Rohde D, et al. Systematic review and meta-analysis of the prevalence of cognitive impairment no dementia in the first year post-stroke. European Stroke Journal. 2019 Jun 1;4(2):160–71.

124. Craig L, Hoo ZL, Yan TZ, Wardlaw J, Quinn TJ. Prevalence of dementia in ischaemic or mixed stroke populations: systematic review and meta-analysis. 2022 Feb 1 [cited 2025 Oct 19]; Available from: https://jnnp.bmj.com/content/93/2/180

125. James BD, Bennett DA, Boyle PA, Leurgans S, Schneider JA. Dementia From Alzheimer Disease and Mixed Pathologies in the Oldest Old. JAMA. 2012 May 2;307(17):1798–800.

126. Nichols E, Merrick R, Hay SI, Himali D, Himali JJ, Hunter S, et al. The prevalence, correlation, and co-occurrence of neuropathology in old age: harmonisation of 12 measures across six community-based autopsy studies of dementia. [cited 2025 Oct 2]; Available from: https://www.thelancet.com/journals/lanhl/article/PIIS2666-7568(23)00019-3/fulltext

127. Nichols E, Szoeke CEI, Vollset SE, Abbasi N, Abd-Allah F, Abdela J, et al. Global, regional, and national burden of Alzheimer's disease and other dementias, 1990–2016: a systematic analysis for the Global Burden

of Disease Study 2016. [cited 2025 Oct 19]; Available from: https://www.thelancet.com/journals/laneur/article/PIIS1474-4422(18)30403-4/fulltext

128. Wolters FJ, Chibnik LB, Waziry R, Anderson R, Berr C, Beiser A, et al. Twenty-seven-year time trends in dementia incidence in Europe and the United States. Neurology. 2020 Aug 4;95(5):e519–31.
129. Livingston G, Huntley J, Liu KY, Costafreda SG, Selbæk G, Alladi S, et al. Dementia prevention, intervention, and care: 2024 report of the Lancet standing Commission. [cited 2025 Aug 6]; Available from: https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(24)01296-0/abstract?_fsi=0YzsSmVe
130. Bai W, Chen P, Cai H, Zhang Q, Su Z, Cheung T, et al. Worldwide prevalence of mild cognitive impairment among community dwellers aged 50 years and older: a meta-analysis and systematic review of epidemiology studies. Age Ageing. 2022 Aug 2;51(8):afac173.
131. Qin J, He Z, Wu L, Wang W, Lin Q, Lin Y, et al. Prevalence of mild cognitive impairment in patients with hypertension: a systematic review and meta-analysis. Hypertension Research. 2021 Oct;44(10):1251–60.
132. Röhr S, Pabst A, Baber R, Engel C, Glaesmer H, Hinz A, et al. Social determinants and lifestyle factors for brain health: implications for risk reduction of cognitive decline and dementia. Scientific Reports. 2022 Jul 28;12(1):12965.
133. Wendeln AC, Degenhardt K, Kaurani L, Gertig M, Ulas T, Jain G, et al. Innate immune memory in the brain shapes neurological disease hallmarks. Nature. 2018 Apr;556(7701):332–8.
134. Ferretti MT, Iulita MF, Cavedo E, Chiesa PA, Schumacher Dimech A, Santuccione Chadha A, et al. Sex differences in Alzheimer disease — the gateway to precision medicine. Nature Reviews Neurology. 2018 Aug;14(8):457–69.
135. Salzman T, Sarquis-Adamson Y, Son S, Montero-Odasso M, Fraser S. Associations of Multidomain Interventions With Improvements in Cognition in Mild Cognitive Impairment: A Systematic Review and Meta-analysis. JAMA Netw Open. 2022 May 3;5(5):e226744.
136. Wimo A, Handels R, Antikainen R, Eriksdotter M, Jönsson L, Knapp M, et al. Dementia prevention: The potential long-term cost-effectiveness of the FINGER prevention program. Alzheimer's & Dementia. 2023;19(3):999–1008.
137. Ngandu T, Lehtisalo J, Solomon A, Levälahti E, Ahtiluoto S, Antikainen R, et al. A 2 year multidomain intervention of diet, exercise, cognitive training, and vascular risk monitoring versus control to prevent cognitive decline in at-risk elderly people (FINGER): a randomised controlled trial. The Lancet. 2015 Jun 6;385(9984):2255–63.
138. Kivipelto M, Mangialasche F, Ngandu T. Lifestyle interventions to prevent cognitive impairment, dementia and Alzheimer disease. Nat Rev Neurol. 2018 Nov;14(11):653–66.
139. Andrieu S, Guyonnet S, Coley N, Cantet C, Bonnefoy M, Bordes S, et al. Effect of long-term omega 3 polyunsaturated fatty acid supplementation with or without multidomain intervention on cognitive function in elderly adults with memory complaints (MAPT): a randomised, placebo-controlled trial. The Lancet Neurology. 2017 May 1;16(5):377–89.
140. Solomon A, Turunen H, Ngandu T, Peltonen M, Levälahti E, Helisalmi S, et al. Effect of the Apolipoprotein E Genotype on Cognitive Change During a Multidomain Lifestyle Intervention: A Subgroup Analysis of a Randomized Clinical Trial. JAMA Neurol. 2018 Apr 1;75(4):462.
141. Delrieu J, Payoux P, Carrié I, Cantet C, Weiner M, Vellas B, et al. Multidomain intervention and/or omega-3 in nondemented elderly subjects according to amyloid status. Alzheimer's & Dementia. 2019 Nov 1;15(11):1392–401.
142. Frisoni GB, Altomare D, Ribaldi F, Villain N, Brayne C, Mukadam N, et al. Dementia prevention in memory clinics: recommendations from the European task force for brain health services. The Lancet Regional Health - Europe. 2023 Mar;26:100576.
143. Chhetri JK, de Souto Barreto P, Cantet C, Pothier K, Cesari M, Andrieu S, et al. Effects of a 3-Year Multi-Domain Intervention with or without Omega-3 Supplementation on Cognitive Functions in Older Subjects with Increased CAIDE Dementia Scores. Journal of Alzheimer's Disease. 2018 Jun 8;64(1):71–8.
144. Fink HA, Jutkowitz E, McCarten JR, Hemmy LS, Butler M, Davila H, et al. Pharmacologic Interventions to Prevent Cognitive Decline, Mild Cognitive Impairment, and Clinical Alzheimer-Type Dementia. Ann Intern Med. 2018 Jan 2;168(1):39–51.
145. Profyri E, Leung P, Huntley J, Orgeta V. Effectiveness of treatments for people living with severe dementia: A systematic review and meta-analysis of randomised controlled clinical trials. Ageing Research Reviews. 2022 Dec 1;82:101758.

146. Budd Haeberlein S, Aisen PS, Barkhof F, Chalkias S, Chen T, Cohen S, et al. Two Randomized Phase 3 Studies of Aducanumab in Early Alzheimer's Disease. The Journal of Prevention of Alzheimer's Disease. 2022 Mar 18;9(2):197–210.

147. Panoutsopoulou K, Wheeler E. Key Concepts in Genetic Epidemiology. In: Evangelou E, editor. Genetic Epidemiology: Methods and Protocols [Internet]. New York, NY: Springer New York; 2018. p. 7–24. Available from: https://doi.org/10.1007/978-1-4939-7868-7_2

148. Fallin MD, Duggal P, Beaty TH. Genetic Epidemiology and Public Health: The Evolution From Theory to Technology. Am J Epidemiol. 2016 Mar 1;183(5):387–93.

149. Ertekin-Taner N. Genetics of Alzheimer's Disease: A Centennial Review. Neurologic Clinics. 2007 Aug 1;25(3):611–67.

150. Rich TA, Woodson AH, Litton J, Arun B. Hereditary breast cancer syndromes and genetic testing. Journal of Surgical Oncology. 2015;111(1):66–80.

151. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The Sequence of the Human Genome. Science. 2001 Feb 16;291(5507):1304–51.

152. Ozaki K, Ohnishi Y, Iida A, Sekine A, Yamada R, Tsunoda T, et al. Functional SNPs in the lymphotoxin-α gene that are associated with susceptibility to myocardial infarction. Nat Genet. 2002 Dec;32(4):650–4.

153. Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K. A comprehensive review of genetic association studies. Genet Med. 2002 Mar;4(2):45–61.

154. Dudbridge F. Power and Predictive Accuracy of Polygenic Risk Scores. PLOS Genetics. 2013 Mar 21;9(3):e1003348.

155. Uffelmann E, Huang QQ, Munung NS, de Vries J, Okada Y, Martin AR, et al. Genome-wide association studies. Nat Rev Methods Primers. 2021 Aug 26;1(1):59.

156. Schuster SC. Next-generation sequencing transforms today's biology. Nat Methods. 2008 Jan;5(1):16–8.

157. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, et al. A global reference for human genetic variation. Nature. 2015 Oct;526(7571):68–74.

158. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nature Genetics. 2016 Oct;48(10):1279–83.

159. Taliun D, Harris DN, Kessler MD, Carlson J, Szpiech ZA, Torres R, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. Nature. 2021 Feb;590(7845):290–9.

160. Hellwege JN, Keaton JM, Giri A, Gao X, Velez Edwards DR, Edwards TL. Population Stratification in Genetic Association Studies. Current Protocols in Human Genetics. 2017;95(1):1.22.1-1.22.23.

161. Kraft P, Zeggini E, Ioannidis JPA. Replication in Genome-Wide Association Studies. Statistical Science. 2009 Nov;24(4):561–73.

162. Sollis E, Mosaku A, Abid A, Buniello A, Cerezo M, Gil L, et al. The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. Nucleic Acids Res. 2023 Jan 6;51(D1):D977–85.

163. Tenesa A, Haley CS. The heritability of human disease: estimation, uses and abuses. Nat Rev Genet. 2013 Feb;14(2):139–49.

164. Young AI. Solving the missing heritability problem. [cited 2025 Oct 22]; Available from: https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1008222

165. Boomsma D, Busjahn A, Peltonen L. Classical twin studies and beyond. Nat Rev Genet. 2002 Nov;3(11):872–82.

166. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009 Oct;461(7265):747–53.

167. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, et al. Common SNPs explain a large proportion of the heritability for human height. Nat Genet. 2010 Jul;42(7):565–9.

168. Tang M, Wang T, Zhang X. A review of SNP heritability estimation methods. Brief Bioinform. 2022 May 1;23(3):bbac067.

169. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Patterson N, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet. 2015 Mar;47(3):291–5.

170. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet. 2015 Nov;47(11):1236–41.

171. Wray NR, Yang J, Hayes BJ, Price AL, Goddard ME, Visscher PM. Pitfalls of predicting complex traits from SNPs. Nat Rev Genet. 2013 Jul;14(7):507–15.

172. Lee SH, DeCandia TR, Ripke S, Yang J, Sullivan PF, Goddard ME, et al. Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. Nat Genet. 2012 Mar;44(3):247–50.

173. Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, et al. Missing heritability and strategies for finding the underlying causes of complex disease. Nat Rev Genet. 2010 Jun;11(6):446–50.

174. Wainschtein P, Jain D, Zheng Z, Cupples LA, Shadyab AH, McKnight B, et al. Assessing the contribution of rare variants to complex trait heritability from whole-genome sequence data. Nat Genet. 2022 Mar;54(3):263–73.

175. Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk to disease from genome-wide association studies. Genome Res. 2007 Jan 10;17(10):1520–8.

176. Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature. 2009 Aug;460(7256):748–52.

177. Choi SW, Mak TSH, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. Nat Protoc. 2020 Sep;15(9):2759–72.

178. Evans LM, Tahmasbi R, Vrieze SI, Abecasis GR, Das S, Gazal S, et al. Comparison of methods that use whole genome data to estimate the heritability and genetic architecture of complex traits. Nat Genet. 2018 May;50(5):737–45.

179. Karch CM, Cruchaga C, Goate AM. Alzheimer's Disease Genetics: From the Bench to the Clinic. Neuron. 2014 Jul 2;83(1):11–26.

180. Guerreiro RJ, Gustafson DR, Hardy J. The genetic architecture of Alzheimer's disease: beyond APP, PSENs and APOE. Neurobiology of Aging. 2012 Mar 1;33(3):437–56.

181. Debette S, Markus HS. Stroke Genetics: Discovery, Insight Into Mechanisms, and Clinical Perspectives. Circulation Research. 2022 Apr 15;130(8):1095–111.

182. Hara K, Shiga A, Fukutake T, Nozaki H, Miyashita A, Yokoseki A, et al. Association of HTRA1 Mutations and Familial Ischemic Cerebral Small-Vessel Disease. New England Journal of Medicine. 2009 Apr 23;360(17):1729–39.

183. Gould DB, Phalan FC, Mil SE van, Sundberg JP, Vahedi K, Massin P, et al. Role of COL4A1 in Small-Vessel Disease and Hemorrhagic Stroke. New England Journal of Medicine. 2006 Apr 6;354(14):1489–96.

184. Bugiani M, Kevelam SH, Bakels HS, Waisfisz Q, Ceuterick-de Groote C, Niessen HWM, et al. Cathepsin A–related arteriopathy with strokes and leukoencephalopathy (CARASAL). Neurology. 2016 Oct 25;87(17):1777–86.

185. Kavanagh D, Spitzer D, Kothari P, Shaikh A, Liszewski MK, Richards A, et al. New roles for the major human 3′-5′ exonuclease TREX1 in human disease. Cell Cycle. 2008 Jun 15;7(12):1718–25.

186. Falcone GJ, Woo D. Genetics of Spontaneous Intracerebral Hemorrhage. Stroke. 2017 Dec;48(12):3420–4.

187. Traylor M, Persyn E, Tomppo L, Klasson S, Abedi V, Bakker MK, et al. Genetic basis of lacunar stroke: a pooled analysis of individual patient data and genome-wide association studies. The Lancet Neurology. 2021 May 1;20(5):351–61.

188. Atwood LD, Wolf PA, Heard-Costa NL, Massaro JM, Beiser A, D'Agostino RB, et al. Genetic Variation in White Matter Hyperintensity Volume in the Framingham Study. Stroke. 2004 Jul;35(7):1609–13.

189. Markus HS. Unravelling the Genetics of Ischaemic Stroke. PLOS Medicine. 2010 Mar 2;7(3):e1000225.

190. Turner ST, Fornage M, Jack CR, Mosley TH, Kardia SLR, Boerwinkle E, et al. Genomic Susceptibility Loci for Brain Atrophy in Hypertensive Sibships From the GENOA Study. Hypertension. 2005 Apr;45(4):793–8.

191. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A Tool for Genome-wide Complex Trait Analysis. The American Journal of Human Genetics. 2011 Jan 7;88(1):76–82.

192. Duperron MG, Tzourio C, Sargurupremraj M, Mazoyer B, Soumaré A, Schilling S, et al. Burden of Dilated Perivascular Spaces, an Emerging Marker of Cerebral Small Vessel Disease, Is Highly Heritable. Stroke. 2018 Feb;49(2):282–7.

193. Bevan S, Traylor M, Adib-Samii P, Malik R, Paul NLM, Jackson C, et al. Genetic Heritability of Ischemic Stroke and the Contribution of Previously Reported Candidate Gene and Genomewide Associations. Stroke. 2012 Dec;43(12):3161–7.

194. Devan WJ, Falcone GJ, Anderson CD, Jagiella JM, Schmidt H, Hansen BM, et al. Heritability Estimates Identify a Substantial Genetic Contribution to Risk and Outcome of Intracerebral Hemorrhage. Stroke. 2013 Jun;44(6):1578–83.

195. Andrews SJ, Renton AE, Fulton-Howard B, Podlesny-Drabiniok A, Marcora E, Goate AM. The complex genetic architecture of Alzheimer's disease: novel insights and future directions. eBioMedicine [Internet]. 2023 Apr 1 [cited 2025 Oct 24];90. Available from: https://www.thelancet.com/journals/ebiom/article/PIIS2352-3964(23)00076-2/fulltext

196. Bertram L, Lill CM, Tanzi RE. The Genetics of Alzheimer Disease: Back to the Future. [cited 2025 Oct 24]; Available from: https://www.cell.com/neuron/abstract/S0896-6273(10)00837-8

197. The Mega Vascular Cognitive Impairment and Dementia (MEGAVCID) consortium. A genome-wide association meta-analysis of all-cause and vascular dementia. Alzheimer's & Dementia. 2024;20(9):5973–95.

198. Persyn E, Hanscombe KB, Howson JMM, Lewis CM, Traylor M, Markus HS. Genome-wide association study of MRI markers of cerebral small vessel disease in 42,310 participants. Nature Communications. 2020 May 1;11(1):2175.

199. Schilling S, DeStefano AL, Sachdev PS, Choi SH, Mather KA, DeCarli CD, et al. APOE genotype and MRI markers of cerebrovascular disease. Neurology. 2013 Jul 16;81(3):292–300.

200. Markus HS, Schmidt R. Genetics of Vascular Cognitive Impairment. Stroke. 2019 Mar;50(3):765–72.

201. Karlsson IK, Escott-Price V, Gatz M, Hardy J, Pedersen NL, Shoai M, et al. Measuring heritable contributions to Alzheimer's disease: polygenic risk score analysis with twins. Brain Commun. 2022 Feb 1;4(1):fcab308.

202. Gatz M, Reynolds CA, Fratiglioni L, Johansson B, Mortimer JA, Berg S, et al. Role of Genes and Environments for Explaining Alzheimer Disease. Arch Gen Psychiatry. 2006 Feb 1;63(2):168–74.

203. Zhang Q, Sidorenko J, Couvy-Duchesne B, Marioni RE, Wright MJ, Goate AM, et al. Risk prediction of late-onset Alzheimer's disease implies an oligogenic architecture. Nat Commun. 2020 Sep 23;11(1):4799.

204. Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet. 2013 Dec;45(12):1452–8.

205. Chakkarai S, Le Grand Q, Shaoxuan LW, Debette S, Sargurupremraj M. Unravelling the genetic architecture of cerebral small vessel disease in the context of stroke. J Cereb Blood Flow Metab. 2025 Aug 6;0271678X251362977.

206. Escott-Price V, Hardy J. Genome-wide association studies for Alzheimer's disease: bigger is not always better. Brain Communications. 2022 Jun 1;4(3):fcac125.

207. Speed D, Kaphle A, Balding DJ. SNP-based heritability and selection analyses: Improved models and new results. BioEssays. 2022;44(5):2100170.

208. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet. 2018 Sep;50(9):1219–24.

209. Abul-Husn NS, Manickam K, Jones LK, Wright EA, Hartzel DN, Gonzaga-Jauregui C, et al. Genetic identification of familial hypercholesterolemia within a single U.S. health care system. Science. 2016 Dec 23;354(6319):aaf7000.

210. Rutten-Jacobs LC, Larsson SC, Malik R, Rannikmäe K, Consortium M, Consortium ISG, et al. Genetic risk, incident stroke, and the benefits of adhering to a healthy lifestyle: cohort study of 306 473 UK Biobank participants. BMJ. 2018 Oct 24;363:k4168.

211. Abraham G, Malik R, Yonova-Doing E, Salim A, Wang T, Danesh J, et al. Genomic risk score offers predictive performance comparable to clinical risk factors for ischaemic stroke. Nature Communications. 2019 Dec 20;10(1):5819.

212. Myserlis EP, Georgakis MK, Demel SL, Sekar P, Chung J, Malik R, et al. A Genomic Risk Score Identifies Individuals at High Risk for Intracerebral Hemorrhage. Stroke. 2023 Apr;54(4):973–82.

213. Mayerhofer E, Parodi L, Prapiadou S, Malik R, Rosand J, Georgakis MK, et al. Genetic Risk Score Improves Risk Stratification for Anticoagulation-Related Intracerebral Hemorrhage. Stroke. 2023 Mar;54(3):791–9.

214. Clocchiatti-Tuozzo S, Rivier CA, Renedo D, Huo S, de Havenon A, Hawkes MA, et al. APOE ε4 and Risk of Intracranial Hemorrhage in Patients With Atrial Fibrillation Taking Apixaban. JAMA Neurol. 2025 Aug 1;82(8):808–16.

215. Marston NA, Patel PN, Kamanu FK, Nordio F, Melloni GM, Roselli C, et al. Clinical Application of a Novel Genetic Risk Score for Ischemic Stroke in Patients With Cardiometabolic Disease. Circulation. 2021 Feb 2;143(5):470–8.

216. Stocker H, Möllers T, Perna L, Brenner H. The genetic risk of Alzheimer's disease beyond APOE ε4: systematic review of Alzheimer's genetic risk scores. Transl Psychiatry. 2018 Aug 24;8(1):1–9.

217. Leonenko G, Baker E, Stevenson-Hoare J, Sierksma A, Fiers M, Williams J, et al. Identifying individuals with high risk of Alzheimer's disease using polygenic risk scores. Nat Commun. 2021 Jul 23;12(1):4506.

218. Ware EB, Faul JD, Mitchell CM, Bakulski KM. Considering the APOE locus in Alzheimer's disease polygenic scores in the Health and Retirement Study: a longitudinal panel study. BMC Medical Genomics. 2020 Nov 3;13(1):164.

219. de Rojas I, Moreno-Grau S, Tesi N, Grenier-Boley B, Andrade V, Jansen IE, et al. Common variants in Alzheimer's disease and risk stratification by polygenic risk scores. Nat Commun. 2021 Jun 7;12(1):3417.

220. Riaz M, Huq A, Ryan J, Orchard SG, Tiller J, Lockery J, et al. Effect of APOE and a polygenic risk score on incident dementia and cognitive decline in a healthy older population. Aging Cell. 2021;20(6):e13384.

221. Najar J, van der Lee SJ, Joas E, Wetterberg H, Hardy J, Guerreiro R, et al. Polygenic risk scores for Alzheimer's disease are related to dementia risk in APOE ε4 negatives. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring. 2021;13(1):e12142.

222. Clark K, Fu W, Liu CL, Ho PC, Wang H, Lee WP, et al. The prediction of Alzheimer's disease through multi-trait genetic modeling. Front Aging Neurosci. 2023 Jul 27;15:1168638.

223. Leonenko G, Shoai M, Bellou E, Sims R, Williams J, Hardy J, et al. Genetic risk for alzheimer disease is distinct from genetic risk for amyloid deposition. Annals of Neurology. 2019;86(3):427–35.

224. Harrison JR, Foley SF, Baker E, Bracher-Smith M, Holmans P, Stergiakouli E, et al. Pathway-specific polygenic scores for Alzheimer's disease are associated with changes in brain structure in younger and older adults. Brain Communications. 2023 Oct 1;5(5):fcad229.

225. Stocker H, Perna L, Weigl K, Möllers T, Schöttker B, Thomsen H, et al. Prediction of clinical diagnosis of Alzheimer's disease, vascular, mixed, and all-cause dementia by a polygenic risk score and APOE status in a community-based cohort prospectively followed over 17 years. Mol Psychiatry. 2021 Oct;26(10):5812–22.

226. Daunt P, Ballard CG, Creese B, Davidson G, Hardy J, Oshota O, et al. Polygenic Risk Scoring is an Effective Approach to Predict Those Individuals Most Likely to Decline Cognitively Due to Alzheimer's Disease. J Prev Alzheimers Dis. 2021 Jan 1;8(1):78–83.

227. Wray NR, Lin T, Austin J, McGrath JJ, Hickie IB, Murray GK, et al. From Basic Science to Clinical Application of Polygenic Risk Scores: A Primer. JAMA Psychiatry. 2021 Jan 1;78(1):101–9.

228. Weale ME, Riveros-Mckay F, Selzam S, Seth P, Moore R, Tarran WA, et al. Validation of an Integrated Risk Tool, Including Polygenic Risk Score, for Atherosclerotic Cardiovascular Disease in Multiple Ethnicities and Ancestries. The American Journal of Cardiology. 2021 Jun 1;148:157–64.

229. Samani NJ, Beeston E, Greengrass C, Riveros-McKay F, Debiec R, Lawday D, et al. Polygenic risk score adds to a clinical risk score in the prediction of cardiovascular disease in a clinical setting. European Heart Journal. 2024 Sep 7;45(34):3152–60.

230. Marston NA, Pirruccello JP, Melloni GEM, Koyama S, Kamanu FK, Weng LC, et al. Predictive Utility of a Coronary Artery Disease Polygenic Risk Score in Primary Prevention. JAMA Cardiology. 2023 Feb 1;8(2):130–7.

231. Mega JL, Stitziel NO, Smith JG, Chasman DI, Caulfield MJ, Devlin JJ, et al. Genetic risk, coronary heart disease events, and the clinical benefit of statin therapy: an analysis of primary and secondary prevention trials. [cited 2025 Mar 21]; Available from: https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(14)61730-X/abstract?elsca4=Public+Health%7CInfectious+Diseases%7CHealth+Policy%7CInternal%2FFamily+Medicine%7CGeneral+Surgery%7CLancet

232. Damask A, Steg PG, Schwartz GG, Szarek M, Hagström E, Badimon L, et al. Patients With High Genome-Wide Polygenic Risk Scores for Coronary Artery Disease May Receive Greater Clinical Benefit From Alirocumab Treatment in the ODYSSEY OUTCOMES Trial. Circulation. 2020 Feb 25;141(8):624–36.

233. Marston NA, Kamanu FK, Nordio F, Gurmu Y, Roselli C, Sever PS, et al. Predicting Benefit From Evolocumab Therapy in Patients With Atherosclerotic Disease Using a Genetic Risk Score. Circulation. 2020 Feb 25;141(8):616–23.

234. Raghavan S, Jablonski K, Delahanty LM, Maruthur NM, Leong A, Franks PW, et al. Interaction of diabetes genetic risk and successful lifestyle modification in the Diabetes Prevention Programme. Diabetes, Obesity and Metabolism. 2021;23(4):1030–40.

235. Trajanoska K, Bhérer C, Taliun D, Zhou S, Richards JB, Mooser V. From target discovery to clinical drug development with human genetics. Nature. 2023 Aug;620(7975):737–45.

236. Fahed AC, Philippakis AA, Khera AV. The potential of polygenic scores to improve cost and efficiency of clinical trials. Nat Commun. 2022 May 25;13(1):2922.

237. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Current clinical use of polygenic scores will risk exacerbating health disparities. Nat Genet. 2019 Apr;51(4):584–91.

238. Wand H, Lambert SA, Tamburro C, Iacocca MA, O'Sullivan JW, Sillari C, et al. Improving reporting standards for polygenic scores in risk prediction studies. Nature. 2021 Mar;591(7849):211–9.

239. Baker E, Escott-Price V. Polygenic Risk Scores in Alzheimer's Disease: Current Applications and Future Directions. Front Digit Health [Internet]. 2020 Aug 11 [cited 2025 Oct 27];2. Available from: https://www.frontiersin.org/journals/digital-health/articles/10.3389/fdgth.2020.00014/full

240. Adeyemo A, Balaconis MK, Darnes DR, Fatumo S, Granados Moreno P, Hodonsky CJ, et al. Responsible use of polygenic risk scores in the clinic: potential benefits, risks and gaps. Nat Med. 2021 Nov;27(11):1876–84.

241. Fuat A, Adlen E, Monane M, Coll R, Groves S, Little E, et al. A polygenic risk score added to a QRISK®2 cardiovascular disease risk calculator demonstrated robust clinical acceptance and clinical utility in the primary care setting. Eur J Prev Cardiol. 2024 Apr 1;31(6):716–22.

242. Natarajan P, Young R, Stitziel NO, Padmanabhan S, Baber U, Mehran R, et al. Polygenic Risk Score Identifies Subgroup With Higher Burden of Atherosclerosis and Greater Relative Benefit From Statin Therapy in the Primary Prevention Setting. Circulation. 2017 May 30;135(22):2091–101.

243. Kiflen M, Le A, Mao S, Lali R, Narula S, Xie F, et al. Cost-Effectiveness of Polygenic Risk Scores to Guide Statin Therapy for Cardiovascular Disease Prevention. Circulation: Genomic and Precision Medicine. 2022 Oct;15(5):e003423.

244. Vascular Factors and Risk of Dementia: Design of the Three-City Study and Baseline Characteristics of the Study Population. Neuroepidemiology. 2003;22(6):316–25.

245. Samieri C, Féart C, Proust-Lima C, Peuchant E, Tzourio C, Stapf C, et al. Olive oil consumption, plasma oleic acid, and stroke incidence. Neurology. 2011 Aug 2;77(5):418–25.

246. World Health Organization. International Statistical Classification of Diseases and Related Health Problems 10th revision [Internet]. [cited 2024 Dec 4]. Available from: https://icd.who.int/browse10/2016/en

247. Godin O, Dufouil C, Maillard P, Delcroix N, Mazoyer B, Crivello F, et al. White Matter Lesions as a Predictor of Depression in the Elderly: The 3C-Dijon Study. Biological Psychiatry. 2008 Apr 1;63(7):663–9.

248. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. Nat Protoc. 2010 Sep;5(9):1564–73.

249. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. Nat Genet. 2016 Oct;48(10):1284–7.

250. on behalf of the MEMENTO cohort Study Group, Dufouil C, Dubois B, Vellas B, Pasquier F, Blanc F, et al. Cognitive and imaging markers in non-demented subjects attending a memory clinic: study design and baseline findings of the MEMENTO cohort. Alz Res Therapy. 2017 Dec;9(1):67.

251. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease. Neurology. 1984 Jul;34(7):939–939.

252. Operto G, Chupin M, Batrancourt B, Habert MO, Colliot O, Benali H, et al. CATI: A Large Distributed Infrastructure for the Neuroimaging of Cohorts. Neuroinform. 2016 Jul 1;14(3):253–64.

253. Gunter J, Thostenson K, Borowski B, Reid R, Arani A, Bernstein M, et al. ADNI-3 MRI protocol. Alzheimer's & Dementia. 2017;13(7):P104–5.

254. Montagni I, Guichard E, Kurth T. Association of screen time with self-perceived attention problems and hyperactivity levels in French students: a cross-sectional study. BMJ Open. 2016 Feb 1;6(2):e009089.

255. Le Grand Q, Satizabal CL, Sargurupremraj M, Mishra A, Soumaré A, Laurent A, et al. Genomic Studies Across the Lifespan Point to Early Mechanisms Determining Subcortical Volumes. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging. 2022 Jun 1;7(6):616–28.

256. Tsuchida A, Laurent A, Crivello F, Petit L, Joliot M, Pepe A, et al. The MRi-Share database: brain imaging in a cross-sectional cohort of 1870 university students. Brain Struct Funct. 2021 Sep 1;226(7):2057–85.

257. Alfaro-Almagro F, Jenkinson M, Bangerter NK, Andersson JLR, Griffanti L, Douaud G, et al. Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. NeuroImage. 2018 Feb 1;166:400–24.

258. Schulman S, Kearon C. Definition of major bleeding in clinical investigations of antihemostatic medicinal products in non-surgical patients. Journal of Thrombosis and Haemostasis. 2005 Apr 1;3(4):692–4.

259. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. [cited 2025 Oct 6]; Available from: https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001779

260. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018 Oct;562(7726):203–9.

261. null null. The "All of Us" Research Program. New England Journal of Medicine. 2019 Aug 15;381(7):668–76.

262. Genomic data in the All of Us Research Program. Nature. 2024;627(8003):340–6.

263. All of Us Research Program: Genomic Research Data Quality Report. [Internet]. National Institutes of Health; [cited 2024 Feb 4]. Available from: https://support.researchallofus.org/hc/en-us/article_attachments/19370367115796

264. Thiébaut ACM, Bénichou J. Choice of time-scale in Cox's model analysis of epidemiologic cohort data: a simulation study. Statistics in Medicine. 2004;23(24):3803–20.

265. Lamarca R, Alonso J, Gómez G, Muñoz Á. Left-truncated Data With Age as Time Scale: An Alternative for Survival Analysis in the Elderly Population. J Gerontol A Biol Sci Med Sci. 1998 Sep 1;53A(5):M337–43.

266. Turkson AJ, Ayiah-Mensah F, Nimoh V. Handling Censoring and Censored Data in Survival Analysis: A Standalone Systematic Literature Review. International Journal of Mathematics and Mathematical Sciences. 2021;2021(1):9307475.

267. Stegherr R, Allignol A, Meister R, Schaefer C, Beyersmann J. Estimating cumulative incidence functions in competing risks data with dependent left-truncation. Statistics in Medicine. 2020;39(4):481–93.

268. Gerds TA, Ohlendorff JS, Ozenne B. riskRegression: Risk Regression Models and Prediction Scores for Survival Analysis with Competing Risks [Internet]. 2023. Available from: https://CRAN.R-project.org/package=riskRegression

269. Km L, Rm E, Aa A. Censoring issues in survival analysis. PubMed [Internet]. [cited 2025 Oct 13]; Available from: https://pubmed.ncbi.nlm.nih.gov/9143713/

270. Radke BR. A demonstration of interval-censored survival analysis. Preventive Veterinary Medicine. 2003 Jun 26;59(4):241–56.

271. Austin PC, Lee DS, Fine JP. Introduction to the Analysis of Survival Data in the Presence of Competing Risks. Circulation. 2016 Feb 9;133(6):601–9.

272. Leffondré K, Touraine C, Helmer C, Joly P. Interval-censored time-to-event and competing risk with death: is the illness-death model more accurate than the Cox model? Int J Epidemiol. 2013 Aug 1;42(4):1177–86.

273. Andersen PK, Geskus RB, de Witte T, Putter H. Competing risks in epidemiology: possibilities and pitfalls. Int J Epidemiol. 2012 Jun 1;41(3):861–70.

274. Kaplan EL, Meier P. Nonparametric Estimation from Incomplete Observations. Journal of the American Statistical Association. 1958;53(282):457–81.

275. Cox DR. Regression Models and Life-Tables. Royal Statistical Society Journal Series B: Methodological. 1972 Jan 1;34(2):187–202.

276. Lacny S, Wilson T, Clement F, Roberts DJ, Faris P, Ghali WA, et al. Kaplan–Meier survival analysis overestimates cumulative incidence of health-related events in competing risk settings: a meta-analysis. Journal of Clinical Epidemiology. 2018 Jan 1;93:25–35.

277. Waller M, Mishra GD, Dobson AJ. A comparison of cause-specific and competing risk models to assess risk factors for dementia. 2020 Jan 1 [cited 2025 Oct 10]; Available from: https://www.degruyterbrill.com/document/doi/10.1515/em-2019-0036/html

278. Weuve J, Proust-Lima C, Power MC, Gross AL, Hofer SM, Thiébaut R, et al. Guidelines for reporting methodological challenges and evaluating potential bias in dementia research. Alzheimer's & Dementia. 2015;11(9):1098–109.

279. Schuster NA, Hoogendijk EO, Kok AAL, Twisk JWR, Heymans MW. Ignoring competing events in the analysis of survival data may lead to biased results: a nonmathematical illustration of competing risk analysis. Journal of Clinical Epidemiology. 2020 Jun 1;122:42–8.

280. Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. Journal of the American Statistical Association. 1999 Jun 1;94(446):496–509.

281. Lau B, Cole SR, Gange SJ. Competing Risk Regression Models for Epidemiologic Data. Am J Epidemiol. 2009 Jul 15;170(2):244–56.

282. Austin PC, Fine JP. Practical recommendations for reporting Fine-Gray model analyses for competing risk data. Statistics in Medicine. 2017;36(27):4391–400.

283. Chatterjee N, Shi J, García-Closas M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. Nat Rev Genet. 2016 Jul;17(7):392–406.

284. Dudbridge F. Polygenic Epidemiology. Genetic Epidemiology. 2016;40(4):268–72.

285. Privé F, Vilhjálmsson BJ, Aschard H, Blum MGB. Making the Most of Clumping and Thresholding for Polygenic Scores. Am J Hum Genet. 2019 Dec 5;105(6):1213–21.

286. Vilhjálmsson BJ, Yang J, Finucane HK, Gusev A, Lindström S, Ripke S, et al. Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. The American Journal of Human Genetics. 2015 Oct;97(4):576–92.

287. Ge T, Chen CY, Ni Y, Feng YCA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. Nat Commun. 2019 Apr 16;10(1):1776.

288. Privé F, Arbel J, Vilhjálmsson BJ. LDpred2: better, faster, stronger. Bioinformatics. 2021 Apr 1;36(22–23):5424–31.

289. Ni G, Zeng J, Revez JA, Wang Y, Zheng Z, Ge T, et al. A Comparison of Ten Polygenic Score Methods for Psychiatric Disorders Applied Across Multiple Cohorts. [cited 2025 Oct 7]; Available from: https://www.biologicalpsychiatryjournal.com/article/S0006-3223(21)01263-4/abstract

290. Ruan Y, Lin YF, Feng YCA, Chen CY, Lam M, Guo Z, et al. Improving polygenic prediction in ancestrally diverse populations. Nature Genetics. 2022 May;54(5):573–80.

291. Lambert SA, Gil L, Jupp S, Ritchie SC, Xu Y, Buniello A, et al. The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation. Nature Genetics. 2021 Apr;53(4):420–5.

292. Truong B, Hull LE, Ruan Y, Huang QQ, Hornsby W, Martin H, et al. Integrative polygenic risk score improves the prediction accuracy of complex traits and diseases. [cited 2025 Oct 8]; Available from: https://www.cell.com/cell-genomics/abstract/S2666-979X(24)00065-X

293. Inouye M, Abraham G, Nelson CP, Wood AM, Sweeting MJ, Dudbridge F, et al. Genomic Risk Prediction of Coronary Artery Disease in 480,000 Adults. Journal of the American College of Cardiology. 2018 Oct 16;72(16):1883–93.

294. Albiñana C, Zhu Z, Schork AJ, Ingason A, Aschard H, Brikell I, et al. Multi-PGS enhances polygenic prediction by combining 937 polygenic scores. Nature Communications. 2023 Aug 5;14(1):4702.

295. Maier RM, Zhu Z, Lee SH, Trzaskowski M, Ruderfer DM, Stahl EA, et al. Improving genetic prediction by leveraging genetic correlations among human diseases and traits. Nature Communications. 2018 Mar 7;9(1):989.

296. Krapohl E, Patel H, Newhouse S, Curtis CJ, von Stumm S, Dale PS, et al. Multi-polygenic score approach to trait prediction. Molecular Psychiatry. 2018 May;23(5):1368–74.

297. Li C, Yang C, Gelernter J, Zhao H. Improving genetic risk prediction by leveraging pleiotropy. Human Genetics. 2013 Dec 13;133(5):639–50.

298. Zou H, Hastie T. Regularization and Variable Selection Via the Elastic Net. Journal of the Royal Statistical Society Series B: Statistical Methodology. 2005 Apr 1;67(2):301–20.

299. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining [Internet]. New York, NY, USA: Association for Computing Machinery; 2016 [cited 2025 Oct 8]. p. 785–94. (KDD '16). Available from: https://dl.acm.org/doi/10.1145/2939672.2939785

300. Darst BF, Koscik RL, Racine AM, Oh JM, Krause RA, Carlsson CM, et al. Pathway-Specific Polygenic Risk Scores as Predictors of Amyloid-β Deposition and Cognitive Function in a Sample at Increased Risk for Alzheimer's Disease. Journal of Alzheimer's Disease. 2017 Jan 1;55(2):473–84.

301. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research. 2000 Jan 1;28(1):27–30.

302. Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, et al. The reactome pathway knowledgebase. Nucleic Acids Research. 2020 Jan 8;48(D1):D498–503.

303. Nishimura D. BioCarta. Biotech Software & Internet Report. 2001 Jun;2(3):117–20.

304. Schaefer CF, Anthony K, Krupa S, Buchoff J, Day M, Hannay T, et al. PID: the Pathway Interaction Database. Nucleic Acids Research. 2009 Jan 1;37(suppl_1):D674–9.

305. Leeuw CA de, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. PLoS Comput Biol [Internet]. 2015 [cited 2025 Mar 15]; Available from: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004219

306. Yu G, Tam CHT, Lim CKP, Shi M, Lau ESH, Ozaki R, et al. Type 2 diabetes pathway-specific polygenic risk scores elucidate heterogeneity in clinical presentation, disease progression and diabetic complications in 18,217 Chinese individuals with type 2 diabetes. Diabetologia. 2024 Nov 12;68(3):602–14.

307. Smith K, Deutsch AJ, McGrail C, Kim H, Hsu S, Huerta-Chagoya A, et al. Multi-ancestry polygenic mechanisms of type 2 diabetes. Nature Medicine. 2024 Apr;30(4):1065–74.

308. Suzuki K, Hatzikotoulas K, Southam L, Taylor HJ, Yin X, Lorenz KM, et al. Genetic drivers of heterogeneity in type 2 diabetes pathophysiology. Nature. 2024 Mar;627(8003):347–57.

309. Maj C, Salvi E, Citterio L, Borisov O, Simonini M, Glorioso V, et al. Dissecting the Polygenic Basis of Primary Hypertension: Identification of Key Pathway-Specific Components. Front Cardiovasc Med [Internet]. 2022 Feb 16 [cited 2025 Feb 12];9. Available from: https://www.frontiersin.org/journals/cardiovascular-medicine/articles/10.3389/fcvm.2022.814502/full

310. Bandres-Ciga S, Saez-Atienzar S, Kim JJ, Makarious MB, Faghri F, Diez-Fairen M, et al. Large-scale pathway specific polygenic risk and transcriptomic community network analysis identifies novel functional pathways in Parkinson disease. Acta Neuropathol. 2020 Sep 1;140(3):341–58.

311. Shams H, Shao X, Santaniello A, Kirkish G, Harroud A, Ma Q, et al. Polygenic risk score association with multiple sclerosis susceptibility and phenotype in Europeans. Brain. 2023 Feb 1;146(2):645–56.

312. Lorenzini L, Collij LE, Tesi N, Vilor-Tejedor N, Ingala S, Blennow K, et al. Alzheimer's disease genetic pathways impact cerebrospinal fluid biomarkers and imaging endophenotypes in non-demented individuals. Alzheimer's & Dementia. 2024 Sep;20(9):6146–60.

313. Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data. GigaScience. 2019 Jul 1;8(7):giz082.

314. Choi SW, García-González J, Ruan Y, Wu HM, Porras C, Johnson J, et al. PRSet: Pathway-based polygenic risk score analyses and software. PLOS Genetics. 2023 févr;19(2):e1010624.

315. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh PR, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nature Genetics. 2015 Nov;47(11):1228–35.

316. Zhou G, Zhao H. A fast and robust Bayesian nonparametric method for prediction of complex traits using summary statistics. PLoS Genet. 2021 Jul 26;17(7):e1009697.

317. Steyerberg EW, Bleeker SE, Moll HA, Grobbee DE, Moons KGM. Internal and external validation of predictive models: A simulation study of bias and precision in small samples. Journal of Clinical Epidemiology. 2003 May 1;56(5):441–7.

318. Steyerberg EW, Harrell FE. Prediction models need appropriate internal, internal–external, and external validation. [cited 2025 Oct 13]; Available from: https://www.jclinepi.com/article/S0895-4356(15)00175-4/abstract

319. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the Performance of Prediction Models: A Framework for Traditional and Novel Measures. Epidemiology [Internet]. 2010;21(1). Available from: https://journals.lww.com/epidem/fulltext/2010/01000/assessing_the_performance_of_prediction_models__a.22.aspx

320. Hao L. Development of a clinical polygenic risk score assay and reporting workflow. Nature Medicine.

321. Frisoni GB, Molinuevo JL, Altomare D, Carrera E, Barkhof F, Berkhof J, et al. Precision prevention of Alzheimer's and other dementias: Anticipating future needs in the control of risk factors and implementation of disease-modifying therapies. Alzheimer's & Dementia. 2020;16(10):1457–68.

322. Klarin D, Natarajan P. Clinical utility of polygenic risk scores for coronary artery disease. Nature Reviews Cardiology. 2022 May;19(5):291–301.

323. Roselli C, Surakka I, Olesen MS, Sveinbjornsson G, Marston NA, Choi SH, et al. Meta-analysis of genome-wide associations and polygenic risk prediction for atrial fibrillation in more than 180,000 cases. Nat Genet. 2025 Mar;57(3):539–47.

324. Visseren FLJ, Mach F, Smulders YM, Carballo D, Koskinas KC, Bäck M, et al. 2021 ESC Guidelines on cardiovascular disease prevention in clinical practice: Developed by the Task Force for cardiovascular disease prevention in clinical practice with representatives of the European Society of Cardiology and 12 medical societies With the special contribution of the European Association of Preventive Cardiology (EAPC). Eur Heart J. 2021 Sep 7;42(34):3227–337.
325. Schunkert H, Di Angelantonio E, Inouye M, Patel RS, Ripatti S, Widen E, et al. Clinical utility and implementation of polygenic risk scores for predicting cardiovascular disease: A clinical consensus statement of the ESC Council on Cardiovascular Genomics, the ESC Cardiovascular Risk Collaboration, and the European Association of Preventive Cardiology. European Heart Journal. 2025 Apr 14;46(15):1372–83.
326. Tsuchida A, Boutinaud P, Verrecchia V, Tzourio C, Debette S, Joliot M. Early detection of white matter hyperintensities using SHIVA-WMH detector. Human Brain Mapping. 2024;45(1):e26548.
327. Wang Z, Chen Q, Chen J, Yang N, Zheng K. Risk factors of cerebral small vessel disease: A systematic and meta-analysis. Medicine. 2021 Dec 23;100(51):e28229.
328. Harshfield EL, Sands CJ, Tuladhar AM, de Leeuw FE, Lewis MR, Markus HS. Metabolomic profiling in small vessel disease identifies multiple associations with disease severity. Brain. 2022 Jul 1;145(7):2461–71.
329. Ye S, Feng K, Zeng G, Cai J, Liang L, Chen J, et al. Frontiers | Deep and periventricular white matter hyperintensities exhibit differential metabolic profiles in arteriosclerotic cerebral small vessel disease: an untargeted metabolomics study. [cited 2025 Sep 16]; Available from: https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2025.1607242/full
330. Gupta A, Thirugnanam K, Thamilarasan M, Mohieldin AM, Zedan HT, Prabhudesai S, et al. Cilia proteins are biomarkers of altered flow in the vasculature. JCI Insight. 7(6):e151813.
331. Pala R, Jamal M, Alshammari Q, Nauli SM. The Roles of Primary Cilia in Cardiovascular Diseases. Cells. 2018 Nov 27;7(12):233.
332. Luu VZ, Chowdhury B, Al-Omran M, Hess DA, Verma S. Role of endothelial primary cilia as fluid mechanosensors on vascular health. Atherosclerosis. 2018 Aug 1;275:196–204.
333. Mohieldin AM, Zubayer HSM, Omran AJA, Saternos HC, Zarban AA, Nauli SM, et al. Vascular Endothelial Primary Cilia: Mechanosensation and Hypertension. 2016 Apr 1 [cited 2025 Sep 17]; Available from: https://www.benthamdirect.com/content/journals/chyr/10.2174/1573402111666150630140615
334. Ma N, Zhou J. Frontiers | Functions of Endothelial Cilia in the Regulation of Vascular Barriers. [cited 2025 Sep 18]; Available from: https://www.frontiersin.org/journals/cell-and-developmental-biology/articles/10.3389/fcell.2020.00626/full
335. Neumann JT, Riaz M, Bakshi A, Polekhina G, Thao LTP, Nelson MR, et al. Predictive Performance of a Polygenic Risk Score for Incident Ischemic Stroke in a Healthy Older Population. Stroke. 2021 Sep;52(9):2882–91.
336. Bellou E, Kim W, Leonenko G, Tao F, Simmonds E, Wu Y, et al. Benchmarking Alzheimer's disease prediction: personalised risk assessment using polygenic risk scores across various methodologies and genome-wide studies. Alz Res Therapy. 2025 Jan 6;17(1):6.
337. Neuffer J, Wagner M, Moreno E, Le Grand Q, Mishra A, Trégouët DA, et al. Association of LIfestyle for BRAin health risk score (LIBRA) and genetic susceptibility with incident dementia and cognitive decline. Alzheimer's & Dementia. 2024;20(6):4250–9.
338. Nicolas A, Sherva R, Grenier-Boley B, Kim Y, Kikuchi M, Timsina J, et al. Transferability of European-derived Alzheimer's disease polygenic risk scores across multiancestry populations. Nat Genet. 2025 Jul;57(7):1598–610.
339. Kikuchi M, Miyashita A, Hara N, Kasuga K, Saito Y, Murayama S, et al. Polygenic effects on the risk of Alzheimer's disease in the Japanese population. Alzheimer's Research & Therapy. 2024 Feb 27;16(1):45.
340. Dalmasso MC, de Rojas I, Olivar N, Muchnik C, Angel B, Gloger S, et al. The first genome-wide association study in the Argentinian and Chilean populations identifies shared genetics with Europeans in Alzheimer's disease. Alzheimer's & Dementia. 2024;20(2):1298–308.
341. Roberts E, Flaum N, Evans DG. Clinical implementation of polygenic risk scores. Eur J Hum Genet. 2025 Sep 29;1–3.
342. Carver T, Hartley S, Lee A, Cunningham AP, Archer S, Babb de Villiers C, et al. CanRisk Tool—A Web Interface for the Prediction of Breast and Ovarian Cancer Risk and the Likelihood of Carrying Genetic Pathogenic Variants. Cancer Epidemiol Biomarkers Prev. 2021 Mar 5;30(3):469–73.

343. Riveros-Mckay F, Weale ME, Moore R, Selzam S, Krapohl E, Sivley RM, et al. Integrated Polygenic Tool Substantially Enhances Coronary Artery Disease Prediction. Circulation: Genomic and Precision Medicine. 2021 Apr;14(2):e003304.

344. JÜRISSON M. The Pragmatic EE-PRS Trial Assessing Polygenic Risk Driven Statin Therapy for Cardiovascular Disease Prevention [Internet]. clinicaltrials.gov; 2025 Mar [cited 2025 Oct 29]. Report No.: NCT06820086. Available from: https://clinicaltrials.gov/study/NCT06820086

345. University of Bern. Polygenic Risk Score to Optimize Primary Prevention in Intermediate Risk Population (PERSONAL) [Internet]. clinicaltrials.gov; 2025 Jul [cited 2025 Oct 29]. Report No.: NCT07039123. Available from: https://clinicaltrials.gov/study/NCT07039123

346. Sun L, Pennells L, Kaptoge S, Nelson CP, Ritchie SC, Abraham G, et al. Polygenic risk scores in cardiovascular risk prediction: A cohort study and modelling analyses. PLoS Med. 2021 Jan;18(1):e1003498.

347. Koriath CAM, Kenny J, Ryan NS, Rohrer JD, Schott JM, Houlden H, et al. Genetic testing in dementia — utility and clinical strategies. Nat Rev Neurol. 2021 Jan;17(1):23–36.

348. Kivimäki M, Livingston G, Singh-Manoux A, Mars N, Lindbohm JV, Pentti J, et al. Estimating Dementia Risk Using Multifactorial Prediction Models. JAMA Netw Open. 2023 Jun 13;6(6):e2318132.

349. You J, Zhang YR, Wang HF, Yang M, Feng JF, Yu JT, et al. Development of a novel dementia risk prediction model in the general population: A large, longitudinal, population-based machine-learning study. [cited 2025 Oct 29]; Available from: https://www.thelancet.com/journals/eclinm/article/PIIS2589-5370(22)00395-9/fulltext

350. Fowler NR, Perkins AJ, Gao S, Sachs GA, Boustani MA. Risks and Benefits of Screening for Dementia in Primary Care: The Indiana University Cognitive Health Outcomes Investigation of the Comparative Effectiveness of Dementia Screening (IU CHOICE)Trial. Journal of the American Geriatrics Society. 2020;68(3):535–43.

351. Thabtah F, Peebles D, Retzler J, Hathurusingha C. Dementia medical screening using mobile applications: A systematic review with a new mapping model. Journal of Biomedical Informatics. 2020 Nov 1;111:103573.

352. Keshavan A, Pannee J, Karikari TK, Rodriguez JL, Ashton NJ, Nicholas JM, et al. Population-based blood screening for preclinical Alzheimer's disease in a British birth cohort at age 70. Brain. 2021 Feb 1;144(2):434–49.

353. Chaudhury S, Brookes KJ, Patel T, Fallows A, Guetta-Baranes T, Turton JC, et al. Alzheimer's disease polygenic risk score as a predictor of conversion from mild-cognitive impairment. Transl Psychiatry. 2019 May 24;9(1):154.

354. Pyun JM, Park YH, Lee KJ, Kim S, Saykin AJ, Nho K, et al. Predictability of polygenic risk score for progression to dementia and its interaction with APOE ε4 in mild cognitive impairment. Transl Neurodegener. 2021 Aug 31;10(1):32.

355. de Rojas I, Hulsman M, Tesi N, van Spaendonk RML, van der Schaar J, Dijkstra JIR, et al. Genetic testing of common and rare variants in dementia patients from a memory clinic. Alzheimer's Research & Therapy. 2025 Oct 14;17(1):225.

356. Ebenau JL, van der Lee SJ, Hulsman M, Tesi N, Jansen IE, Verberk IMW, et al. Risk of dementia in APOE ε4 carriers is mitigated by a polygenic risk score. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring. 2021;13(1):e12229.

357. Koohi F, Harshfield EL, Gill D, Ge W, Burgess S, Markus HS. Optimizing treatment of cardiovascular risk factors in cerebral small vessel disease using genetics. Brain. 2025 Jun 1;148(6):1936–49.

358. Lan LF, Zheng L, Yang X, Ji XT, Fan YH, Zeng JS. Peroxisome Proliferator-activated Receptor-γ Agonist Pioglitazone Ameliorates White Matter Lesion and Cognitive Impairment in Hypertensive Rats. CNS Neuroscience & Therapeutics. 2015;21(5):410–6.

359. Shirzadi Z, Schultz AP, Loghmani N, Yang HS, Ford J, Yaari R, et al. Independent effects of white matter lesion volume and APOE ε4 on ARIA-H in A4 Study. Alzheimer's & Dementia. 2025;21(10):e70751.

360. Salloway S, Wojtowicz J, Voyle N, Lane CA, Klein G, Lyons M, et al. Amyloid-Related Imaging Abnormalities (ARIA) in Clinical Trials of Gantenerumab in Early Alzheimer Disease. JAMA Neurol. 2025 Jan 1;82(1):19–29.

361. Ritchie M, Sajjadi SA, Grill JD. Apolipoprotein E Genetic Testing in a New Age of Alzheimer Disease Clinical Practice. Neurology Clinical Practice. 2024 Apr;14(2):e200230.

362. Van Der Lee SJ, Hulsman M, Van Spaendonk R, Van Der Schaar J, Dijkstra J, Tesi N, et al. Prevalence of Pathogenic Variants and Eligibility Criteria for Genetic Testing in Patients Who Visit a Memory Clinic. Neurology. 2025 Feb 25;104(4):e210273.

363. Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomström-Lundqvist C, et al. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): The Task Force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) Developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. European Heart Journal. 2020 Aug;42(5):373–498.

364. Kim J, Kim D, Kim D, Park BE, Kang TS, Lim SH, et al. Polygenic Risk and Cardiovascular Event Risk in Patients With Atrial Fibrillation With Low to Intermediate Stroke Risk. Journal of the American Heart Association [Internet]. 2025 Apr 1 [cited 2025 Jul 31]; Available from: https://www.ahajournals.org/doi/10.1161/JAHA.124.037727

365. O'Sullivan JW, Shcherbina A, Justesen JM, Turakhia M, Perez M, Wand H, et al. Combining Clinical and Polygenic Risk Improves Stroke Prediction Among Individuals With Atrial Fibrillation. Circulation: Genomic and Precision Medicine. 2021 Jun;14(3):e003168.

366. Attelind S, Eriksson N, Wadelius M, Hallberg P. Genome-wide association study of direct oral anticoagulants and their relation to bleeding. Eur J Clin Pharmacol. 2025 Mar 21;1–13.

367. Paré G, Eriksson N, Lehr T, Connolly S, Eikelboom J, Ezekowitz MD, et al. Genetic Determinants of Dabigatran Plasma Levels and Their Relation to Bleeding. Circulation. 2013 Apr 2;127(13):1404–12.

368. Zhai S, Zhang H, Mehrotra DV, Shen J. Pharmacogenomics polygenic risk score for drug response prediction using PRS-PGx methods. Nature Communications. 2022 Sep 8;13(1):5278.

369. Georgakis MK, Anderson CD. Polygenic Risk Scores in the Clinic: A Case for Stroke Survivors. Neurology. 2023 Apr 11;100(15):693–5.

370. Dudbridge F, Allen RJ, Sheehan NA, Schmidt AF, Lee JC, Jenkins RG, et al. Adjustment for index event bias in genome-wide association studies of subsequent events. Nature Communications. 2019 Apr 5;10(1):1561.

371. Debette S, Caro I, Western D, Namba S, Sun N, Kawaguchi S, et al. Proteogenomics in cerebrospinal fluid and plasma reveals new biological fingerprint of cerebral small vessel disease. Res Sq. 2024 Jul 2;rs.3.rs-4535534.

372. Wang Y chao, Zhu H hang, He LC, Yao Y ting, Zhang L, Xue X li, et al. Proteome Profiling of Serum Reveals Pathological Mechanisms and Biomarker Candidates for Cerebral Small Vessel Disease. Transl Stroke Res. 2025 Oct 1;16(5):1606–20.

373. Wu C, Li J, Yan J, Li Y, Ya J, Zhao Q, et al. Identification of Shared Genes between Cerebral Small Vessel Disease and Alzheimer's Disease through Multi-Omics Analyses. Neuropsychiatric Sciences and Molecular Biology [Internet]. 2025 Oct 11 [cited 2025 Oct 31];2(1). Available from: https://journalnsmb.com/index.php/nsmb/article/view/16137

374. Mao H, Xu M, Wang H, Liu Y, Wang F, Gao Q, et al. Transcriptional patterns of brain structural abnormalities in CSVD-related cognitive impairment. Front Aging Neurosci [Internet]. 2024 Nov 29 [cited 2025 Oct 31];16. Available from: https://www.frontiersin.org/journals/aging-neuroscience/articles/10.3389/fnagi.2024.1503806/full

375. Song Y, Zhou X, Zhao H, Zhao W, Sun Z, Zhu J, et al. Characterizing the role of the microbiota-gut-brain axis in cerebral small vessel disease: An integrative multi-omics study. NeuroImage. 2024 Dec 1;303:120918.

376. Reus LM, Boltz T, Francia M, Bot M, Ramesh N, Koromina M, et al. Quantitative trait loci mapping of circulating metabolites in cerebrospinal fluid to uncover biological mechanisms involved in brain-related phenotypes. Mol Psychiatry. 2025 Aug;30(8):3478–90.

377. Arteaga-Reyes C, Sen D, Rudilosso S, Jouvent E, Hervé D, Lindgren AG, et al. Time to consider health services dedicated for adults living with cerebral small vessel disease: Report of a ESO scientific seminar. European Stroke Journal. 2025 Sep 1;10(3):675–81.

378. Brunham LR. Polygenic Modulation of Monogenic Diseases: Familial Hypercholesterolemia as the Exemplar. Curr Atheroscler Rep. 2025 Aug 13;27(1):80.

379. Sherva R, Zhang R, Sahelijo N, Jun G, Anglin T, Chanfreau C, et al. African ancestry GWAS of dementia in a large military cohort identifies significant risk loci. Mol Psychiatry. 2023 Mar;28(3):1293–302.

380. Ge YJ, Chen SD, Wu BS, Zhang YR, Wang J, He XY, et al. Genome-wide meta-analysis identifies ancestry-specific loci for Alzheimer's disease. Alzheimer's & Dementia. 2024;20(9):6243–56.

381. Blue EE, Broome J, Xue D, Kingston H, Chapman NH, Gogarten S, et al. Multi-ancestry meta-analysis identifies genetic modifiers of age-at-onset of Alzheimer's disease at known and novel loci. Alzheimer's & Dementia. 2025;21(9):e70489.

382. Willett JDS, Waqas M, Choi Y, Ngai T, Mullin K, Tanzi RE, et al. Identification of 16 novel Alzheimer's disease loci using multi-ancestry meta-analyses. Alzheimer's & Dementia. 2025;21(2):e14592.

383. Kachuri L, Chatterjee N, Hirbo J, Schaid DJ, Martin I, Kullo IJ, et al. Principles and methods for transferring polygenic risk scores across global populations. Nat Rev Genet. 2024 Jan;25(1):8–25.

384. Xiang R, Kelemen M, Xu Y, Harris LW, Parkinson H, Inouye M, et al. Recent advances in polygenic scores: translation, equitability, methods and FAIR tools. Genome Medicine. 2024 Feb 19;16(1):33.

385. Clifton L, Collister JA, Liu X, Littlejohns TJ, Hunter DJ. Assessing agreement between different polygenic risk scores in the UK Biobank. Sci Rep. 2022 Jul 27;12(1):12812.

386. Pain O, Glanville KP, Hagenaars SP, Selzam S, Fürtjes AE, Gaspar HA, et al. Evaluation of polygenic prediction methodology within a reference-standardized framework. Plagnol V, editor. PLoS Genet. 2021 May 4;17(5):e1009021.

387. Pain O, Al-Chalabi A, Lewis CM. The GenoPred pipeline: a comprehensive and scalable pipeline for polygenic scoring. Bioinformatics. 2024 Oct 1;40(10):btae551.

388. Lennon NJ, Kottyan LC, Kachulis C, Abul-Husn NS, Arias J, Belbin G, et al. Selection, optimization and validation of ten chronic disease polygenic risk scores for clinical implementation in diverse US populations. Nat Med. 2024 Feb;30(2):480–7.

389. Choi SW, Mak TSH, Hoggart CJ, O'Reilly PF. EraSOR: a software tool to eliminate inflation caused by sample overlap in polygenic score analyses. GigaScience. 2023 Jan 1;12:giad043.

390. Ellis CA, Oliver KL, Harris RV, Ottman R, Scheffer IE, Mefford HC, et al. Inflation of polygenic risk scores caused by sample overlap and relatedness: Examples of a major risk of bias. The American Journal of Human Genetics. 2024 Sep 5;111(9):1805–9.

391. Lee YH, Thaweethai T, Sheu YH, Feng YCA, Karlson EW, Ge T, et al. Impact of selection bias on polygenic risk score estimates in healthcare settings. Psychological Medicine. 2023 Nov;53(15):7435–45.

392. van Alten S, Domingue BW, Faul J, Galama T, Marees AT. Reweighting UK Biobank corrects for pervasive selection bias due to volunteering. Int J Epidemiol. 2024 Apr 11;53(3):dyae054.

393. Schoeler T, Speed D, Porcu E, Pirastu N, Pingault JB, Kutalik Z. Participation bias in the UK Biobank distorts genetic associations and downstream analyses. Nat Hum Behav. 2023 Jul;7(7):1216–27.

394. Wu Y, Sun Z, Zheng Q, Miao J, Dorn S, Mukherjee S, et al. Pervasive biases in proxy genome-wide association studies based on parental history of Alzheimer's disease. Nat Genet. 2024 Dec;56(12):2696–703.

395. Escott-Price V, Schmidt KM. Pitfalls of predicting age-related traits by polygenic risk scores. Annals of Human Genetics. 2023;87(5):203–9.

396. Newbury A, Elhussein A, Gürsoy G. Multi-domain rule-based phenotyping algorithms enable improved GWAS signal. npj Digit Med. 2025 Aug 2;8(1):499.

397. Beesley LJ, Fritsche LG, Mukherjee B. An analytic framework for exploring sampling and observation process biases in genome and phenome-wide association studies using electronic health records. Statistics in Medicine. 2020;39(14):1965–79.

398. Herzig AF, Clerget-Darpoux F, Génin E. The False Dawn of Polygenic Risk Scores for Human Disease Prediction. Journal of Personalized Medicine. 2022 Jul 31;12(8):1266.

399. Génin E, Clerget-Darpoux F. Revisiting the Polygenic Additive Liability Model through the Example of Diabetes Mellitus. Hum Hered. 2015 Apr 1;80(4):171–7.

400. Janssens ACJW. Validity of polygenic risk scores: are we measuring what we think we are? Hum Mol Genet. 2019 Nov 21;28(R2):R143–50.

401. Sud A, Horton RH, Hingorani AD, Tzoulaki I, Turnbull C, Houlston RS, et al. Realistic expectations are key to realising the benefits of polygenic scores. BMJ. 2023 Mar 1;380:e073149.

402. Hill WG, Goddard ME, Visscher PM. Data and Theory Point to Mainly Additive Genetic Variance for Complex Traits. PLOS Genetics. 2008 févr;4(2):e1000008.

403. Visscher PM, Wray NR. Concepts and Misconceptions about the Polygenic Additive Model Applied to Disease. Hum Hered. 2016 Sep 1;80(4):165–70.

404. Kullo IJ. Clinical use of polygenic risk scores: current status, barriers and future directions. Nat Rev Genet. 2026 Mar;27(3):246–63.

405. Kullo IJ, Jouni H, Austin EE, Brown SA, Kruisselbrink TM, Isseh IN, et al. Incorporating a Genetic Risk Score Into Coronary Heart Disease Risk Estimates. Circulation. 2016 Mar 22;133(12):1181–8.

406. Slunecka JL, van der Zee MD, Beck JJ, Johnson BN, Finnicum CT, Pool R, et al. Implementation and implications for polygenic risk scores in healthcare. Hum Genomics. 2021 Jul 20;15(1):46.

407. Knoppers BM, Bernier A, Granados Moreno P, Pashayan N. Of Screening, Stratification, and Scores. Journal of Personalized Medicine. 2021 Aug;11(8):736.

408. Ohta R, Tanigawa Y, Suzuki Y, Kellis M, Morishita S. A polygenic score method boosted by non-additive models. Nat Commun. 2024 May 29;15(1):4433.

409. Kelemen M, Xu Y, Jiang T, Zhao JH, Anderson CA, Wallace C, et al. Performance of deep-learning-based approaches to improve polygenic scores. Nat Commun. 2025 Jun 2;16(1):5122.

410. Xu Y, Ritchie SC, Liang Y, Timmers PRHJ, Pietzner M, Lannelongue L, et al. An atlas of genetic scores to predict multi-omic traits. Nature. 2023 Apr;616(7955):123–31.

411. van Hilten A, van Rooij J, Ikram MA, Niessen WJ, Meurs JBJ van, Roshchupkin GV. Phenotype prediction using biologically interpretable neural networks on multi-cohort multi-omics data. npj Syst Biol Appl. 2024 Aug 2;10(1):81.

412. Mostafavi H, Harpak A, Agarwal I, Conley D, Pritchard JK, Przeworski M. Variable prediction accuracy of polygenic scores within an ancestry group. Loos R, Eisen MB, O'Reilly P, editors. eLife. 2020 Jan 30;9:e48376.

413. Argentieri MA, Amin N, Nevado-Holgado AJ, Sproviero W, Collister JA, Keestra SM, et al. Integrating the environmental and genetic architectures of aging and mortality. Nat Med. 2025 Mar;31(3):1016–25.

414. Jayasinghe D, Pham VVH, Beckmann K, Benyamin B, Lee SH. Analyzing complex traits and diseases using GxE PRS: genotype-environment interaction in polygenic risk score models. J Hum Genet. 2025 Nov;70(11):547–56.

415. Gauderman WJ, Fu Y, Queme B, Kawaguchi E, Wang Y, Morrison J, et al. Pathway polygenic risk scores (pPRS) for the analysis of gene-environment interaction. [cited 2025 Oct 8]; Available from: https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1011543

416. Meisel SF, Carere DA, Wardle J, Kalia SS, Moreno TA, Mountain JL, et al. Explaining, not just predicting, drives interest in personal genomics. Genome Medicine. 2015 Aug 1;7(1):74.

417. Hollands GJ, French DP, Griffin SJ, Prevost AT, Sutton S, King S, et al. The impact of communicating genetic risks of disease on risk-reducing health behaviour: systematic review with meta-analysis. BMJ. 2016 Mar 15;352:i1102.

418. Silarova B, Sharp S, Usher-Smith JA, Lucas J, Payne RA, Shefer G, et al. Effect of communicating phenotypic and genetic risk of coronary heart disease alongside web-based lifestyle advice: the INFORM Randomised Controlled Trial. 2019 Jul 1 [cited 2025 Oct 31]; Available from: https://heart.bmj.com/content/105/13/982

419. Halmesvaara O, Vornanen M, Kääriäinen H, Perola M, Kristiansson K, Konttinen H. Psychosocial Effects of Receiving Genome-Wide Polygenic Risk Information Concerning Type 2 Diabetes and Coronary Heart Disease: A Randomized Controlled Trial. Front Genet [Internet]. 2022 May 30 [cited 2025 Oct 31];13. Available from: https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2022.881349/full

420. Muse ED, Chen SF, Liu S, Fernandez B, Schrader B, Molparia B, et al. Impact of polygenic risk communication: an observational mobile application-based coronary artery disease study. npj Digit Med. 2022 Mar 11;5(1):30.

421. Linder JE, Allworth A, Bland HT, Caraballo PJ, Chisholm RL, Clayton EW, et al. Returning integrated genomic risk and clinical recommendations: The eMERGE study. Genetics in Medicine. 2023 Apr 1;25(4):100006.

422. Hull LE, Aday AW, Bui QM, Luzum JA, Muchira JM, Wand H, et al. Direct-to-Consumer Genetic Testing for Cardiovascular Disease: A Scientific Statement From the American Heart Association. Circulation. 2025 Apr 8;151(14):e905–17.

Supplementary Appendices

Appendix 1: Accounting for biases in GWAS ____ *226*

Appendix 2: Modern post-GWAS methods ____ *227*

Fine-mapping ____ **227**

Functional Annotation ____ **227**

Mendelian Randomization ____ **228**

Drug discovery and development ____ **229**

Appendix 3: Data Sources and Quality control procedures for PRS ____ *231*

Target Data QC ____ **231**

Base Data QC ____ **233**

Harmonization ____ **233**

Appendix 4: PRS Methods ____ *234*

Appendix 5: Prediction metrics ____ *237*

Supplementary References ____ *241*

Appendix 6: iPRS-DEM (Article 1) Supplementary Materials ____ *244*

Appendix 7: PRESTIGE-AF (Article 2) Supplementary Materials ____ *284*

Appendix 8: WMH-(ps)PRS (Article 3) Supplementary Material ____ *303*

Appendix 1: Accounting for biases in GWAS

Population stratification refers to systematic differences in genotypes between individuals of diverse ancestry. Indeed, we find substantial differences in allele frequencies throughout the genome across diverse populations, and these differences lead to bias in GWAS (false-positive or false-negative associations) if not appropriately accounted for (1). For example, if a phenotype of interest occurs more frequently in one ancestry group compared to another, then allele frequency differences will be artificially associated with this trait. At a high-level, one needs to account for diverse ancestries and GWAS are traditionally performed in individuals that are of the same ancestry, however, there still often subtle population substructures present within populations of the same ancestry (2). The most common way to account for population stratification even after ancestry matching is to adjust regression models in GWAS by genetic principal components (PCs) constructed using principal component analysis (PCA) (1,3). Briefly, PCA is dimension reduction technique that summarizes the genetic variance across the genotypes of all individuals. These PCAs can be used to visualize genetic relationships between samples and to infer ancestral background when compared against a known reference population, often samples from the 1000 Genomes Project (2). Genetic outliers identified by PCA are often removed. The development of methods that permit cross- or trans-ancestry GWAS is well underway, and have shown to improve translation of genomic insights across ancestries (4).

Relatedness also naturally introduces bias in GWAS by violating the statistical assumption that all samples are independent, and can be either known based on family structure or unknown (i.e. cryptic relatedness). Typically, individuals with high relatedness are removed from a sample but this leads to a loss of statistical power. Mixed models that include a random effect term to account for genetic relatedness based on a genomic relationship matrix are incorporated in more recent GWAS software that more efficiently handle the increased computational burden of these models (3,5). Accounting for these sources of bias is standard procedure for GWAS, and software and protocols have been developed to harmonize GWAS in meta-analysis across multiple studies.

Appendix 2: Modern post-GWAS methods

Fine-mapping

As mentioned, GWAS identifies many indirect associations between SNPs and a phenotype where the SNP is in LD with the casual variant. One key follow-up GWAS analysis called fine-mapping prioritizes a set of variants most likely to be causal within the genomic risk loci through observed patterns in linkage disequilibrium and association statistics (3,6) (**Figure S1**). The set of candidate causal variants identified by fine-mapping can be followed up in further functional analysis. Current state-of-the-art Bayesian methods (7) improve fine-mapping resolution through the use of WGS and new imputation panels, by integrating functional annotations (e.g., gene expression data), and combining ethnically diverse populations to leverage differences in LD patterns (6).

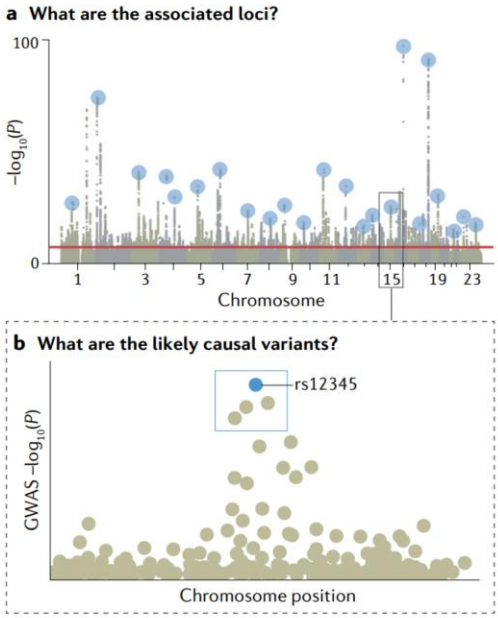


Figure S1. Fine-mapping is used to identify likely casual variants in GWAS.

Functional Annotation

Functional annotation in GWAS primarily focuses on prioritizing the likely affected gene. Typically only 2-3% of GWAS loci are fine-mapped to coding variants where their potential effect on genes can be inferred by existing annotation tools (3,8). However, for the rest of the loci lie outside of coding regions, and have unknown regulatory functions.

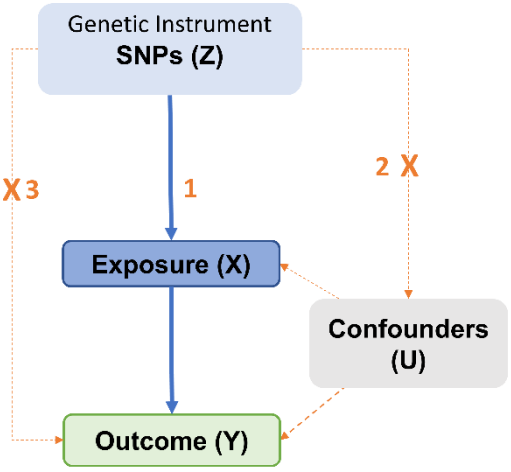
Molecular quantitiative trait loci (molQTLs) is an additional approach that can be used to identify regulatory target genes. These refer to associations between genetic variants and different molecular phenotypes including: expression QTL for RNA expression (eQTL), protein QTL (pQTL) for associations with protein levels typically circulating in blood, or methylation QTLs (meQTL) for variants associated with epigenetic effects. These molQTLs can be evaluated in different tissues or mediums. Public datasets are also becoming available such as GTEx which provides eQTL data across 54 different tissues (9). Leveraging this data can be used to infer the molecular processes mediating the association between a SNP and a phenotype.

Another key follow-up analysis is to perform gene-set and pathway enrichment analysis to determine if genetic loci from GWAS converge on certain biological processes. Gene-sets and pathways are based on biological annotations defined in pathway databases such as, REACTOME (10) or the Kyoto Encyclopedia of Genes and Genomes (KEGG) (11). Based on published literature and peer-review, these databases are meticulously curated in a hierarchical structure that identifies the implication of genes and specifically their protein products in molecular transformations characterizing these pathways (10). A widely used tool for gene-set enrichment is MAGMA (12) which directly outputs association statistics at the pathway-level. Other databases such as Gene Ontology (13) can be used to further investigate the functional importance of identified genes.

Mendelian Randomization

Mendelian randomization (MR) is a key post-GWAS analysis that can be used to assess the casual relationship between an exposure and an outcome by using genetic variants as instrumental variables (14). Instrumental variable analysis is a well-established statistical technique in epidemiology that can allow for casual inference and circumvent issues of confounding if particular statistical assumptions are met. This is possible by exploiting Mendel's 2nd law of independent assortment which asserts that genetic variants are randomly assigned at conception which makes them generally unrelated to confounding environmental factors (14). Essentially, MR can be thought to mimic a randomized control trial which is the gold standard in causal inference. The casual effect is estimated by relating the gene-exposure association and the gene-outcome association.

Box 3 Mendelian Randomization Assumptions



1. **Relevance:** The genetic variant (IV, Z) must be robustly associated with the modifiable exposure of interest (X).
2. **Independence:** The genetic variant (Z) must be independent of the confounding factors (U) that confound the association of the exposure (X) and the outcome (Y).
3. **Exclusion Restriction:** The genetic variant (Z) must be associated with the outcome (Y) only via the exposure of interest (X) and not through other pathways. (no indirect pathways)

In one-sample MR, individual level data from one study is used to assess both the variant-exposure and variant-outcome associations, however, this approach is often limited by

generalizability and statistical power. More popular is two-sample MR (15), which leverages exposure and outcome associations from large-scale GWAS summary statistics without the need for individual-level data. Importantly, MR requires strong statistical assumptions (**Box 3**), primarily revolving around horizontal pleiotropy where genetic variants influences the outcome through indirect pathways other than the exposure (15). Current guidelines urge that MR should be carefully interpreted and findings should be thoroughly investigated through a set of sensitivity analyses (16).

Performing MR is standard procedure in most GWAS publications, especially for endophenotypes thought to precede end-stage disease. In the context of cerebrovascular disease, genetically-predicted WMH have shown evidence for a casual association with vascular risk factors (blood pressure traits, type-2 diabetes, and smoking), and clinical outcomes including small vessel stroke (e.g. lacunar stroke), deep ICH, ischemic stroke more generally, and with AD (17).

Drug discovery and development

GWAS and genomics in general show great potential to accelerate drug development and discovery. Drug development is a lengthy and very costly, with nearly 9/10 drugs failing to gain approval due to insufficient efficacy (18). First GWAS can help to validate the therapeutic hypothesis that a drug target will be beneficial for patients with minimal toxicity (19). Importantly, GWAS provides evidence on humans that is often lacking in animal models, and in observational studies that are subject to confounding and reverse causation. Identifying targets from common variants and functional inference of these GWAS loci can help to prioritize targets. For example, the discovery of a common variant in the HMGCR gene by GWAS, which influences LDL cholesterol levels, provided strong evidence reinforcing HMGCR inhibitors (statins) as successful targets for heart disease prevention (19). Further, mendelian randomization can be used to establish evidence for a casual links between drug targets and disease, and drug target MR methods have been developed to proxy the lifelong consequences of genetic perturbations of drug targets (20). There is also potential for drug repurposing whereby chemical compounds that directly target the protein products of GWAS risk genes could be promising candidates. One example is inhibitors of interleukin-23 (IL-23) signalling, developed for psoriasis, were repurposed for Crohn's Disease based on evidence from common variants in the *IL23R* gene in GWAS (18). Drug discovery efforts are facilitated

by pipelines such as Open Targets (21) and GREP (22) that integrate GWAS risk genes with other molecular datasets and clinical indications.

Appendix 3: Data Sources and Quality control procedures for PRS

PRS relies on two distinct data sources (23):

(1) GWAS summary statistics to extract SNP effect sizes – referred to as **Base Data**.

(2) Genotyped data from an independent cohort—referred to as **Target Data.**

A key requirement for the target data is that it must be **independent** (non-overlapping) of both the **base data**. Failure to ensure independence causes 'over-fitting' which greatly biases or inflates PRS prediction and associations in the target cohort. This is due to a correlation induced between estimated SNP effects and the resulting PRS association leading to spurious inflation, and even minimal overlap between base and target data can result in highly significant associations for non-heritable traits (24). Accounting for sample overlap is an ongoing challenge since many cohorts with genetic data have contributed to large-scale GWAS meta-analyses.

Strict quality control procedures must also be applied to both base and target datasets. It is further recommended to use an external reference panel to standardized quality control of PRS to ensure they are robust, comparable and more generalizable to other cohorts (25). A reference panel is an ancestry-matched individual-level genetic dataset that is used to obtain unbiased estimations of genetic properties like LD and MAF. Importantly, the use of a reference panel ensures that PRS is independent of any properties unique to the target sample such as its population structure, or particular genetic variation in a population arising from non-random mating, and LD or MAF estimates. The 1000 Genomes Project (Phase 3) provides ancestry-specific reference panels that are widely used and considered as a global reference for human genetic variation (26).

Target Data QC

Target data should undergo genetic imputation, a process used to greatly extend coverage cross the genome by estimating genotypes at locations not directly measured by the genotyping array. Genotyping arrays (or chips) target specific, known locations in the DNA and typically only include about 1 million markers. Imputation is performed on essentially every cohort contributing to GWAS meta-analyses, and requires a reference panel, such as from 1000 Genomes Project or the haplotype reference consortium (27), where genetic variation has been extensively characterized. The imputation algorithm compares the patterns of LD observed in segments of an individual's genome with haplotypes (sets of closely linked genetic markers) on the reference panel and infers missing genotypes by observing which unmeasured variants

typically co-occur with measured variants on the reference panel. The output of imputed genotypes are given probabilistically and converted into dosages of the expected number of copies of the effect allele, either 0, 1, or 2. Additionally, the quality of the imputation for each variant is measured by an imputation quality score (INFO score), which ranges between 0 (complete uncertainty) and 1 (complete certainty) (28). This imputation is typically performed on an external server with an automated pipeline, such as the Michigan Imputation Server which utilizes the MiniMac3 algorithm (28).

Prior to imputation, quality control is performed at the SNP and sample (individual) levels on the directly genotyped data in the target cohort (29). This includes filtering out SNPs with:

- **Low MAF** (<1%)
- **Low call rate** (<95%): these SNPs have a high level of missingness across samples are considered poor quality.
- **Deviation from Hardy-Weinberg equilibrium** ($p < 1x10^{-6}$): This indicates whether genotype frequencies differ substantially from what is expected, and deviation could be a source of genotyping error.
- **Ambiguous SNPs**: SNPs where the risk allele cannot be unambiguously identified (e.g., A/T or C/G SNPs) these are also called palindromic SNPs

And individuals with:

- **High Missingness (>95%):** individuals missing a portion of their genetic data
- **High Heterozygosity Rates (≥3SD):** Individuals who deviate too far from the sample's mean heterozygosity rate (proportion of heterozygous genotypes) may be a result of inbreeding or reflect poor sample quality.
- **Sex Discordance:** individuals whose genetically inferred sex does not match their reported sex
- **Divergent ancestry (≥4SD):** individuals identified as outliers for population structure based on principal components of an ancestry-specific reference panel (e.g., 1000G European reference panel)

Once the sample is imputed, an additional SNP-level QC is performed:

- Removal of poorly imputed SNPs (INFO < 0.8)
- Removal of rare imputed SNPs (MAF < 0.01)

After this final step, the target dataset contains only high-quality genotyped data.

Base Data QC

For base data from GWAS meta-analyses these quality control steps have already been performed in each cohort. However, GWAS summary statistics do not provide any individual-level information, and will only sometimes provide imputation quality information (thresholds are often lower for imputation quality in GWAS studies, INFO <0.4). While standardized reporting of GWAS is recommended, often publicly available summary statistics have incomplete information. The following checks should be performed on base data:

- Removal of SNPs with MAF < 0.01 or INFO < 0.8 (if applicable)
- Removal of ambiguous SNPs
- **SNP-based heritability:** The SNP-based heritability (h^2_{SNP}) is a measure of the phenotypic variance that is explained by all the common SNPs included in the GWAS – in fact this represents the upper limit of PRS predictive power. Usually this is estimated by linkage disequilibrium score regression (LDSC) (30), and it is recommend to avoid constructing PRS from GWAS summary statistics with h^2_{SNP}<0.05 (5%) (23).

The MungeSumstats R package (31) provides a thorough pipeline to perform quality control checks on GWAS summary statistics, and is also useful to remove duplicated SNPs, non-biallelic SNPs, non-autosomal SNPs, and to check for data entry errors and harmonize file formats for downstream analyses.

Harmonization

The final step is to ensure SNPs in both the target and base datasets are harmonized, ideally against an ancestry-matched reference panel, so that the effect alleles in the GWAS summary data align correctly with the alleles in the target data. This includes:

- **Strand-Flipping:** Addressing instances where alleles are reported on opposite DNA strands (e.g., A/C in base/target data vs. T/G on the reference panel).
- **Effect Allele Frequency (EAF):** Comparing EAFs between the base/target data and the reference panel serves as a good general sanity check, especially for palindromic SNPs, as differences could indicate population or technical mismatches

Appendix 4: PRS Methods

In C+T three hyper-parameters are tuned – the correlation threshold (r^2) and the genetic distance (w) for clumping, and the p-value threshold (p_t). Typically, in PRS studies both the r^2 (0.1 or 0.2) and w (250 or 500) are pre-determined, and various p_t values are tested ranging from $5x10^{-8}$ (GWS) to 1.0 (32). To note, multiple SNPs can be retained within the same genomic region if the effects are indeed independent, in other words, clumping does not simply retain only the most associated SNP in a region (23). A recent study shows that the tuning of the three hyper-parameters in addition to an imputation quality (INFO) threshold improves prediction compared to only tuning p_t, and recommends performing PRS across a large grid of possible parameters (33). Both INFO and genetic distance were more consistent across different traits whereas r^2 and p_t tended to optimize over a larger range of values. Since imputation information is not always included in GWAS summary statistics, and most traits were not sensitive to genetic distance, here I only vary r^2 and p_t when applying C+T.

C+T is implementable using PLINK software (34) where a set of independent variants can be defined via clumping (--clump function), and filtered based on different p-value thresholds using the(--q-score-range and –score functions. Alternatively, PRSice-2 is a specialized software that efficiently streamlines the C+T method (35), often using PLINK commands and R-scripts. It can efficiently construct PRS using many r^2 and p-value thresholds, and also assess predictive ability of each PRS parameterization in a target sample while adjusting for covariates and producing visualizations. PRSice-2 is my preferred software and is used extensively in this work.

The LDpred method computes a LD-matrix that infers posterior mean effect sizes of each SNP by conditioning on a genetic architecture prior and LD information from an external reference panel (e.g., either the UKBB LD reference panel or 1000 Genomes)[6]. LDpred2 additionally offers a 'sparse' option that learns which SNP-effects are truly zero, and an 'auto' option to automatically tune these parameters which does not require validation data and performs almost as well as a grid-model that takes considerably more computational effort. LDpred2 is implemented using the *bigsnpr* R package (36), and a tutorial is provided by the authors (https://privefl.github.io/bigsnpr/articles/LDpred2.html).

PRS-cs alternatively uses a continuous shrinkage prior (ϕ) on SNP effect sizes is meant to be adaptive to the strength of its association signal in the GWAS and can accommodate for diverse underlying genetic architectures. The main difference with LDpred comes from jointly updating the effect sizes in each LD block in multivariate fashion versus updating effect sizes for each

SNP separately and sequentially – leading to substantial computational improvements and better modeling of local LD patterns. The software also offers an 'automatic' tuning option. PRS-cs is implemented using the 1000 Genomes European sample as an external LD reference (26).

Advanced PRS methods include PRS-CSx (37), BridgePRS (38), SBayesRC (39) which all propose to improve cross-ancestry prediction, and GAUDI has shown to improve prediction in admixed populations (40). Others make use of functional priors to improve prediction (LDpred-funct (41), MegaPRS (42), SBayesRC (39)), are designed to investigate disease heterogeneity through a pathway-based approach (PRSet (43)), or tailored to predict treatment response by integrating pharmacogenetic GWAS (PRS-PGx-Bayes (44)). PRS can also be optimized for single- or cross-trait prediction by integrating and optimizing across multiple PRS models (PRSmix (45), multi-PGS (46)). Sima et al. provide a comprehensive overview of current PRS methods (47).

PRS Tool	Methodology	Data requirements	Software	Use Case	Source
PLINK2	C + T (manual thresholding)	GWAS summary statistics	PLINK framework	Standard PRS	(Chang et al., 2015)
PRSice-2	C + T	GWAS summary statistics	R and C++	Standard PRS	(Choi and O'Reilly, 2019)
LDpred2	Bayesian regression	GWAS summary statistics and LD reference panel	R package: *bigs npR*	Standard PRS	(Privé, Arbel, and Vilhjálmsson, 2021)
PRS-CS	Bayesian regression with continuous shinkage priors	GWAS summary statistics and LD reference panel	Python based	Standard PRS	(Ge et al., 2019)
Lassosum2	Lasso penalized regression	GWAS summary statistics or individual-level data	R package: *bigs npR*	Standard PRS	(Privé, Arbel, Aschard, et al., 2021)
SBayesR	Bayesian multiple regression	GWAS summary statistics	GCTB software	Standard PRS	(Lloyd-Jones et al., 2019)
SBayesRC	Bayesian multiple regression	GWAS summary statistics and LD reference panel, annotation file	R packages	Functional annotation, multi-ancestry	(Zheng et al. 2024)
LDpred-funct	Bayesian regression	GWAS summary statistics	Python package	Functional annotation	(Márquez-Luna et al., 2021)
PRS-CSx	Bayesian regression with shared continuous shrinkage priors	Multiple GWAS summary statistics	Python based	Multi-ancestry	(Ruan et al., 2022)
BridgePRS	Bayesian regression	GWAS summary statistics	R and BASH	Multi-ancestry	(Hoggart et al., 2023)
PRSet	Pathway-based	GWAS summary statistics	R and C++	Pathway-Specific, etiological	(Choi et al., 2023)
GAUDI	Fused lasso	GWAS summary statistics	R, Python, Shell script	Admixed inviduals	(Sun et al., 2024)
Multi-PGS	Lasso penalized regression and/or XGBoost	Existing PRS models or GWAS summary statistics covariates (optional)	R	Multi-trait	(Albinana et al. 2023)
PRSmix (+)	Elastic net penalized regression	Existing PRS models or GWAS summary statistics	R and PLINK2	Single-trait or cross-trait prediction	(Truong et al. 2024)
PRS-PGx-Bayes	Bayesian regression	Two GWAS summary statistics for disease and pharmacogenetics	R package	Drug response	(Zhai et al. 2022)

Table S1. Description of validated PRS methods from standard approaches to advanced approaches with specific use-cases. Adapted from Sima et al. (2024) (47)

Appendix 5: Prediction metrics

In the case of quantitative or continuous traits, we are only concerned with calibration. Studies typically use the phenotypic variance (R^2) explained by PRS, and in a linear regression model, the incremental- or partial-R^2 is used, assessing the added variance explained after accounting for covariates. However, while these metrics explain the explanatory power of PRS, it does not directly measure we assess how close predicted values are to the observed values (i.e., calibration). Preferentially, one would use a metric like the root-mean-squared error,

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \widehat{y_i})^2}{N}}$$

Where:

- N is the number of observations
- $y_i$is the actual value for the *i-th* observation
- $\widehat{y_i}$ is the predicted value for the *i-th* observation

For binary traits, model evaluation involves assessing both discrimination and calibration. Traditionally, discrimination has been measured using the ROC curve displaying the sensitivity (true positive rate) versus 1 – specificity (false positive rate) at different cut-off thresholds to classify subjects as predicted "cases" (individuals who experience the event) or "controls" (event-free) (**Figure S2A**). The ROC curve is summarized by the area under the curve (AUC) which is then used to compare diagnostic accuracy across models. The AUC can directly be interpreted as the probability that a randomly selected case has a larger predicted risk than a randomly selected control. Its values will range from 0 to 1, with an AUC = 1 having perfect discrimination, AUC = 0.5 indicating the model is no better than chance. Generally, over 0.7 is considered acceptable discrimination, while < 0.5 indicates the model is likely inverted.

Calibration can also be evaluated graphically with a calibration plot by plotting observed (proportion of events) versus predicted probabilities (**Figure S2B**). The 45-degree line represents perfect calibration where predicted risks are exactly matching observed outcomes, and the model can be assessed for systematic over or underestimation of risk. Calibration can be numerically summarized by the calibration intercept, where non-zero intercepts indicate over- or-underestimation of risk, and the calibration slope can be assessed in terms of its

deviation from one. The Hosmer-Lemeshow goodness-of-fit test is often used to formally test if a model is well-calibrated (48).

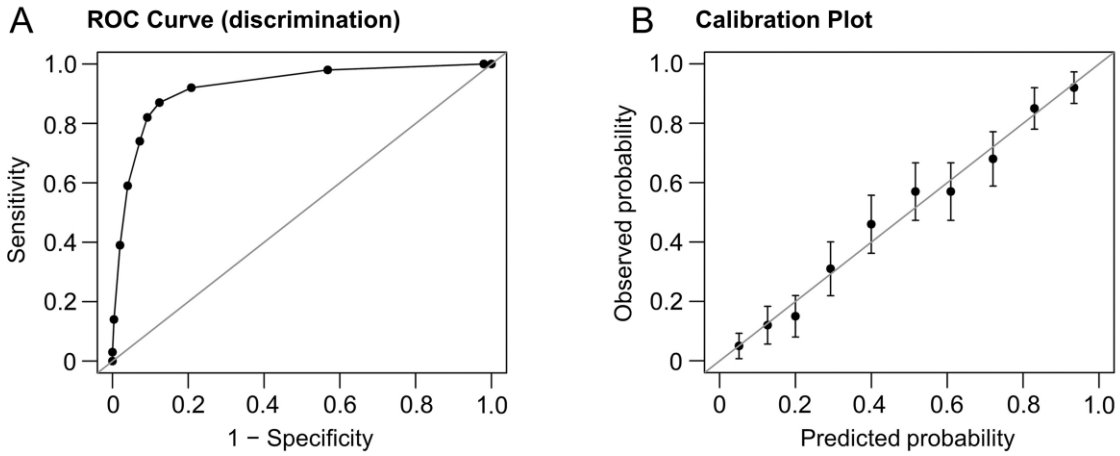


Figure S2: Graphical representation of discriminiation and calibration **(A)** An receiver operating characteristics (ROC) curve wehere X-axis represents the false positive rate (1 – specificity) and the Y-axis represents the true positive rate (sensitivity). **(B)** A calibration plot where the X-axis shows average predicted probability values for each decile, and Y-axis shows corresponding observed probability in each decile. Error bars represent 95% confidence intervals of mean predicted probabilities. The 45-degree line represents perfect calibration.Adapted from Park et al. (2018) (49)

Overall prediction accuracy, combining both discrimination and calibration, is often assessed with the brier score (BS), the average squared difference between the predicted probability of an event and the observed outcome indicator (i.e., 1 for the event and 0 for the non-event) (50).

$$\text{Brier Score} = \frac{1}{N}\sum_{i=1}^{N}(p_i - o_i)^2$$

Where:

- N is the number of observations
- p_i is the predicted probability for the event for observation, i
- o_i is the observed outcome either, 0 or 1

A brier score closer to zero represents better prediction performance. If a model has poor discrimination it will assign similar risk estimates to both cases and controls, meaning that $(p_i - o_i)^2$ will be large, so good discrimination inherently leads to lower brier score. Likewise, poorly calibrated models will systematically over or under-estimate risk, and the Brier Score penalizes larger differences between predicted risk and the binary outcome indicator for each individual. The Brier Score is only a pure error measure, and is not easily interpretable, therefore it is common to scale the brier score based on a null model without covariates.

The null model (without covariates) is a model that estimates the risk equally for all individuals and can in the setting of competing events be estimated by the Aalen-Johansen estimator.13 The scaled Brier score can be interpreted as an R^2 type of measure, representing the amount of prediction error in a null model that is explained by the prediction model.

$$\mathrm{BS}_{\mathrm{scaled}} = 1 - \frac{\sum_{i=1}^{N}(p_i - o_i)^2}{\sum_{i=1}^{N}(\bar{p} - o_i)^2}$$

Where:

- $\bar{p}$ is the average risk or overall observed outcome proportion across all individuals

A 100% 'Scaled' Brier score corresponds to a perfect model, 0% to an ineffective model, and <0% to a harmful model in the sense that the predictions are further away from the observed data than the null model estimating the average risk for each patient.

These different prediction metrics have been adapted for time-to-event outcomes and also in the competing risk setting (51). To start, the concordance index (c-index) is the most widely used metric for evaluating the performance of survival models (52). It is a measure of discrimination, and is exactly equivalent to AUC in a binary setting. For time-to-event outcomes, concordance occurs when a subject who experiences the event earlier in time is assigned a higher predicted risk than the subject who experiences an event later. The c-index is the proportion of concordant pairs out of all comparable pairs.

$$\mathrm{C-Index} = \hat{P}(\mathrm{Concordance}) = \frac{\mathrm{Number\ of\ Concordant\ Pairs} + 0.5(\mathrm{Number\ of\ Indeterminate\ Pairs})}{\mathrm{Number\ of\ Comparable\ Pairs}}.$$

Like the AUC, its values will range from 0.5 to 1. The c-index offers indication of discriminatory performance across the whole study period, however, a number of critics have cautioned against its use (53,54). For instance, at a fixed time horizon, the c-index becomes problematic since it considers the order of event times rather than the order of event status at the prediction horizon, and a time-dependent AUC (AUC_t) is recommended in this case (54,55). Similarly, the Brier Score can be calculated at specific time-horizons (56), and can be scaled in the survival analysis setting as the "Index of Prediction Accuracy" (IPA) (57):

$$IPA = 1 - \frac{BrierScore\ (Prediction\ model)}{BrierScore(NullModel)}$$

Where the null model is an estimator such as Kaplan-Meier (58) (no competing risk) or Aalen-Johansen method (with competing risks) (59). Given the above equation, higher IPA is better, IPA = 1 would indicate a perfect model, IPA = 0 is a useless model, while IPA < 0 is a harmful

model. These metrics also need to account for right censoring in survival data. Briefly, to deal with censoring in survival analysis, both time-dependent AUC and the Brier Score use an inverse probability censoring weighting (IPCW) approach (56) that models the probability of not being censored and re-weights individuals in the dataset. Finally, both the Brier Score and AUC_t can be integrated across multiple time points to offer an overall measure across the study period. For example, the integrated Brier score (IBS) represents global average of the Brier score calculated at the specified time points during follow-up, and can serve as a global metric of prediction performance in a survival analysis setting. *RiskRegression* is a comprehensive R package that allows for easy implementation of all these prediction metrics for survival models (60). I use this package extensively to evaluation prediction models in this work.

Supplementary References

1. Price AL, Zaitlen NA, Reich D, Patterson N. New approaches to population stratification in genome-wide association studies. Nat Rev Genet. 2010 Jul;11(7):459–63.
2. Hellwege JN, Keaton JM, Giri A, Gao X, Velez Edwards DR, Edwards TL. Population Stratification in Genetic Association Studies. Current Protocols in Human Genetics. 2017;95(1):1.22.1-1.22.23.
3. Uffelmann E, Huang QQ, Munung NS, de Vries J, Okada Y, Martin AR, et al. Genome-wide association studies. Nat Rev Methods Primers. 2021 Aug 26;1(1):59.
4. Atkinson EG, Bianchi SB, Ye GY, Martínez-Magaña JJ, Tietz GE, Montalvo-Ortiz JL, et al. Cross-ancestry genomic research: time to close the gap. Neuropsychopharmacol. 2022 Sep;47(10):1737–8.
5. Mbatchou J, Barnard L, Backman J, Marcketta A, Kosmicki JA, Ziyatdinov A, et al. Computationally efficient whole-genome regression for quantitative and binary traits. Nat Genet. 2021 Jul;53(7):1097–103.
6. Schaid DJ, Chen W, Larson NB. From genome-wide associations to candidate causal variants by statistical fine-mapping. Nat Rev Genet. 2018 Aug;19(8):491–504.
7. Wang G, Sarkar A, Carbonetto P, Stephens M. A Simple New Approach to Variable Selection in Regression, with Application to Genetic Fine Mapping. Journal of the Royal Statistical Society Series B: Statistical Methodology. 2020 Dec 1;82(5):1273–300.
8. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010 Sep 1;38(16):e164.
9. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013 Jun;45(6):580–5.
10. Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, et al. The reactome pathway knowledgebase. Nucleic Acids Research. 2020 Jan 8;48(D1):D498–503.
11. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research. 2000 Jan 1;28(1):27–30.
12. Leeuw CA de, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. PLoS Comput Biol [Internet]. 2015 [cited 2025 Mar 15]; Available from: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004219
13. The Gene Ontology Consortium, Aleksander SA, Balhoff J, Carbon S, Cherry JM, Drabkin HJ, et al. The Gene Ontology knowledgebase in 2023. Genetics. 2023 May 2;224(1):iyad031.
14. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?*. Int J Epidemiol. 2003 Feb 1;32(1):1–22.
15. Lawlor DA, Harbord RM, Sterne JAC, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Stat Med. 2008 Apr 15;27(8):1133–63.
16. Burgess S, Davey Smith G, Davies NM, Dudbridge F, Gill D, Glymour MM, et al. Guidelines for performing Mendelian randomization investigations: update for summer 2023. Wellcome Open Res. 2019;4:186.
17. Sargurupremraj M, Suzuki H, Jian X, Sarnowski C, Evans TE, Bis JC, et al. Cerebral small vessel disease genomics and its implications across the lifespan. Nat Commun. 2020 Dec 8;11(1):6285.
18. Trajanoska K, Bhérer C, Taliun D, Zhou S, Richards JB, Mooser V. From target discovery to clinical drug development with human genetics. Nature. 2023 Aug;620(7975):737–45.
19. Plenge RM, Scolnick EM, Altshuler D. Validating therapeutic targets through human genetics. Nat Rev Drug Discov. 2013 Aug;12(8):581–94.
20. Daghlas I, Gill D. Mendelian randomization as a tool to inform drug development using human genetics. Cambridge Prisms: Precision Medicine. 2023 Jan;1:e16.
21. Peat G, Jones W, Nuhn M, Marugán JC, Newell W, Dunham I, et al. The open targets post-GWAS analysis pipeline. Bioinformatics. 2020 May 1;36(9):2936–7.
22. Sakaue S, Okada Y. GREP: genome for REPositioning drugs. Bioinformatics. 2019 Oct 1;35(19):3821–3.

23. Choi SW, Mak TSH, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. Nat Protoc. 2020 Sep;15(9):2759–72.

24. Choi SW, Mak TSH, Hoggart CJ, O'Reilly PF. EraSOR: a software tool to eliminate inflation caused by sample overlap in polygenic score analyses. GigaScience. 2023 Jan 1;12:giad043.

25. Pain O, Glanville KP, Hagenaars SP, Selzam S, Fürtjes AE, Gaspar HA, et al. Evaluation of polygenic prediction methodology within a reference-standardized framework. Plagnol V, editor. PLoS Genet. 2021 May 4;17(5):e1009021.

26. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, et al. A global reference for human genetic variation. Nature. 2015 Oct;526(7571):68–74.

27. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nature Genetics. 2016 Oct;48(10):1279–83.

28. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. Nat Genet. 2016 Oct;48(10):1284–7.

29. Marees AT, de Kluiver H, Stringer S, Vorspan F, Curis E, Marie-Claire C, et al. A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. International Journal of Methods in Psychiatric Research. 2018;27(2):e1608.

30. Vilhjálmsson BJ, Yang J, Finucane HK, Gusev A, Lindström S, Ripke S, et al. Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. The American Journal of Human Genetics. 2015 Oct;97(4):576–92.

31. Murphy AE, Schilder BM, Skene NG. MungeSumstats: a Bioconductor package for the standardization and quality control of many GWAS summary statistics. Bioinformatics. 2021 Dec 7;37(23):4593–6.

32. Wray NR, Lee SH, Mehta D, Vinkhuyzen AAE, Dudbridge F, Middeldorp CM. Research Review: Polygenic methods and their application to psychiatric traits. Journal of Child Psychology and Psychiatry. 2014;55(10):1068–87.

33. Privé F, Vilhjálmsson BJ, Aschard H, Blum MGB. Making the Most of Clumping and Thresholding for Polygenic Scores. Am J Hum Genet. 2019 Dec 5;105(6):1213–21.

34. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaSci. 2015 Dec;4(1):7.

35. Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data. GigaScience. 2019 Jul 1;8(7):giz082.

36. Privé F, Aschard H, Ziyatdinov A, Blum MGB. Efficient analysis of large-scale genome-wide data with two R packages: bigstatsr and bigsnpr. Bioinformatics. 2018 Aug 15;34(16):2781–7.

37. Ruan Y, Lin YF, Feng YCA, Chen CY, Lam M, Guo Z, et al. Improving polygenic prediction in ancestrally diverse populations. Nature Genetics. 2022 May;54(5):573–80.

38. Hoggart CJ, Choi SW, García-González J, Souaiaia T, Preuss M, O'Reilly PF. BridgePRS leverages shared genetic effects across ancestries to increase polygenic risk score portability. Nature Genetics. 2024 Jan;56(1):180–6.

39. Zheng Z, Liu S, Sidorenko J, Wang Y, Lin T, Yengo L, et al. Leveraging functional genomic annotations and genome coverage to improve polygenic prediction of complex traits within and between ancestries. Nature Genetics. 2024 May;56(5):767–77.

40. Sun Q, Rowland BT, Chen J, Mikhaylova AV, Avery C, Peters U, et al. Improving polygenic risk prediction in admixed populations by explicitly modeling ancestral-differential effects via GAUDI. Nature Communications. 2024 Feb 3;15(1):1016.

41. Márquez-Luna C, Gazal S, Loh PR, Kim SS, Furlotte N, Auton A, et al. Incorporating functional priors improves polygenic prediction accuracy in UK Biobank and 23andMe data sets. Nature Communications. 2021 Oct 18;12(1):6052.

42. Zhang Q, Privé F, Vilhjálmsson B, Speed D. Improved genetic prediction of complex traits from individual-level data or summary statistics. Nature Communications. 2021 Jul 7;12(1):4192.

43. Choi SW, García-González J, Ruan Y, Wu HM, Porras C, Johnson J, et al. PRSet: Pathway-based polygenic risk score analyses and software. PLOS Genetics. 2023 févr;19(2):e1010624.

44. Zhai S, Zhang H, Mehrotra DV, Shen J. Pharmacogenomics polygenic risk score for drug response prediction using PRS-PGx methods. Nature Communications. 2022 Sep 8;13(1):5278.

45. Truong B, Hull LE, Ruan Y, Huang QQ, Hornsby W, Martin H, et al. Integrative polygenic risk score improves the prediction accuracy of complex traits and diseases. [cited 2025 Oct 8]; Available from: https://www.cell.com/cell-genomics/abstract/S2666-979X(24)00065-X

46. Albiñana C, Zhu Z, Schork AJ, Ingason A, Aschard H, Brikell I, et al. Multi-PGS enhances polygenic prediction by combining 937 polygenic scores. Nature Communications. 2023 Aug 5;14(1):4702.

47. Ndong Sima CAA, Step K, Swart Y, Schurz H, Uren C, Möller M. Methodologies underpinning polygenic risk scores estimation: a comprehensive overview. Human Genetics. 2024 Oct 19;143(11):1265–80.

48. Nattino G, Pennell ML, Lemeshow S. Assessing the Goodness of Fit of Logistic Regression Models in Large Samples: A Modification of the Hosmer-Lemeshow Test. Biometrics. 2020 Jun 1;76(2):549–60.

49. Park SH, Han K. Methodologic Guide for Evaluating Clinical Performance and Effect of Artificial Intelligence Technology for Medical Diagnosis and Prediction. Radiology. 2018 Mar;286(3):800–9.

50. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the Performance of Prediction Models: A Framework for Traditional and Novel Measures. Epidemiology [Internet]. 2010;21(1). Available from: https://journals.lww.com/epidem/fulltext/2010/01000/assessing_the_performance_of_prediction_models__a.22.aspx

51. Geloven N van, Giardiello D, Bonneville EF, Teece L, Ramspek CL, Smeden M van, et al. Validation of prediction models in the presence of competing risks: a guide through modern methods. 2022 May 24 [cited 2025 Oct 9]; Available from: https://www.bmj.com/content/377/bmj-2021-069249

52. Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. Statistics in Medicine. 2004;23(13):2109–23.

53. Hartman N, Kim S, He K, Kalbfleisch JD. Pitfalls of the concordance index for survival outcomes. Statistics in Medicine. 2023;42(13):2179–90.

54. Blanche P, Kattan MW, Gerds TA. The c-index is not proper for the evaluation of t-year predicted risks. Biostatistics. 2019 Apr 1;20(2):347–57.

55. Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. Statistics in Medicine. 2013;32(30):5381–97.

56. Gerds TA, Schumacher M. Consistent Estimation of the Expected Brier Score in General Survival Models with Right-Censored Event Times. Biometrical Journal. 2006;48(6):1029–40.

57. Kattan MW, Gerds TA. The index of prediction accuracy: an intuitive measure useful for evaluating risk prediction models. Diagn Progn Res. 2018 May 4;2(1):7.

58. Kaplan EL, Meier P. Nonparametric Estimation from Incomplete Observations. Journal of the American Statistical Association. 1958;53(282):457–81.

59. Johansen R. An Empirical Transition Matrix for Non-homogeneous Markov Chains Based on Censored Observations. In 1978 [cited 2025 Oct 9]. Available from: https://www.semanticscholar.org/paper/An-Empirical-Transition-Matrix-for-Non-homogeneous-Johansen/86db6591763c2a7d7285068eb2186d945645e670

60. Gerds TA, Ohlendorff JS, Ozenne B. riskRegression: Risk Regression Models and Prediction Scores for Survival Analysis with Competing Risks [Internet]. 2023. Available from: https://CRAN.R-project.org/package=riskRegression

Appendix 6

Polygenic score integrating neurodegenerative and vascular risk informs dementia risk stratification

Supplement

Table of contents

eMethods 2

eMethods 1. Genotyping and Quality Control 2

eMethods 2. PRS Methods 3

eMethods 3. Construction of iPRS-DEM Model 4

eMethods 4. Clinical risk factor definitions in 3C-Dijon 5

eMethods 5. Fine-Gray Regression Models and Interpretation 5

eMethods 6. Age at onset analysis in 3C-Dijon and Memento 6

eMethods 7. Defining optimal cohort-specific cutoffs to dichotomize iPRS-DEM 6

eMethods 8. Cumulative Incidence of all-cause dementia in genetic risk strata in 3C-Dijon 7

eMethods 9. Risk prediction models in 3C-Dijon 7

eMethods 10. Clinical risk factor definitions and risk prediction models in Memento 7

eMethods 11. Cumulative incidence estimation in Memento 8

Supplementary Figures 9

eFigure 1. Participant exclusion flowchart for (A) Three Cities and (B) Memento. The full 3C-Dijon and Memento analytical sample were used for main association analyses. We performed complete cases analysis for all analyses including clinical risk factors. *In *APOE-ε4* stratified analyses we removed APOE-*ε2/ε4* individuals.†In Memento only low education and family history of dementia were used in prediction models since other risk factors was not associated with increased dementia risk, only 11 individuals were removed from predictive models. 9

eFigure 2. Correlations between all single-trait PRS included in iPRS-DEM. *indicates nominally significant correlations ($p<0.05$) 10

eFigure 3. sPRS associations with cumulative incidence of dementia before and after Fine-Gray Elastic-net Regularization in training set. 11

eFigure 4. Association of AD-PRS with cumulative incidence of dementia in 3C-Dijon across percentile groups. 12

References 13

Data availability 15

Code availability 15

Participating consortia 16

The GIGASTROKE Consortium 16

EADB Consortium 20

Supplementary Tables 26

eMethods

eMethods 1. Genotyping and Quality Control

3C and Memento

Standard quality control procedures were initially conducted on raw genotyped data in both 3C and Memento study populations. This included removing SNPs due to missingness, low minor allele frequency (MAF), non-autosomal location, and deviation from Hardy-Weinberg equilibrium (1). Participants were removed with discordant sex information, excessive missingness, high heterozygosity, or divergent ancestry (1). Additionally, sample pairs were identified with high-relatedness, of which individuals with a lower call rate were removed (2). The raw genotyped data was then imputed via the Michigan server using the Haplotype Reference Consortium (HRC) (3). Post-imputation data were then harmonized to Phase 3 of the 1000 Genomes European subset in order to ensure transferability of PRS into different populations of European ancestry (4,5). In each cohort, SNPs were removed if they had an imputation score of < 0.8 or a MAF < 0.01 in either the target cohort or the reference population, and alleles were flipped to match the reference.

All of Us

Blood samples (15 mL) were collected from consenting participants for genetic analysis, with saliva samples used as an alternative when blood collection was not possible. To date, genetic data have been obtained from 312,945 participants, resulting in 245,388 short-read whole genome sequencing (srWGS) samples that include over 1.03 billion single-nucleotide polymorphisms (SNPs) and insertion/deletion (indel) sites (6,7). Standard quality control (QC) measures were applied to the srWGS data by the AoU team and reported to the Data and Research Center. Uniformity in genomic data was ensured through a consistent library preparation protocol (PCR-Free Lapa HyperPrep), sequencing platform (NovaSeq 6000), and bioinformatics pipeline (DRAGEN v3.4.12). Sex concordance was assessed by comparing self-reported sex assigned at birth with genetic sex determined using DRAGEN; samples with discordant results were excluded, and their sources were investigated for potential sample swaps. Cross-individual contamination rates were evaluated using VerifyBamIDF2 and DRAGEN by measuring read counts at common homozygous alternate SNP sites, while sequencing coverage thresholds, as detailed in prior publications (6), were enforced to maintain adequate statistical power and ensure accurate variant calling. Further QC measures were applied to the joint srWGS call set, including the identification of sample and population outliers, variant filtering using hard thresholds, allele-specific variant quality score recalibration, and the evaluation of sensitivity and precision. A comprehensive description of these additional QC steps is provided elsewhere (6).

GWAS Summary Statistics Quality Control

We performed recommended quality control for PRS construction. This included removing traits with non-significant SNP heritability ($SNP_{h2,}$ p <0.05, estimated by LDSC Regression (8)) harmonizing to the 1000G European reference panel, and re-ran GWAS meta-analyses after exclusion of 3C (stroke, AD) and Memento (AD) to avoid sample overlap with training and validation sets (4,9). In detail, each GWAS summary statistics also underwent similar quality control, removing ambiguous or duplicated variants, variants with a minor allele frequency (MAF) < 0.01 in the GWAS sample or a discordant MAF (>0.2) between the reference and GWAS sample, and lastly, variants with sample size >3SD from the median sample size (if per variant sample size is available) (4). Alleles were also flipped to match the reference at this stage. All quality control and harmonization procedures were conducted using *Plink1.90* software (10,11), the primary software available for management and analysis of SNP data. Additionally, we removed the *APOE* region (chromosome 19: 44.4-46.5Mb) from the AD-GWAS as recommended to optimize AD-PRS (12). Traits that were investigated but removed with non-significant

heritability included large artery stroke (13), brain infarcts (14), cerebral microbleeds (15), longituidinal hippocampal change (16), and age-related hippocampal change (16).

Addressing Sample Overlap

Of note, sample overlap between target and base (GWAS) data is known to bias PRS-trait associations (17,18). 3C had been previously included in the Stroke GWAS (13), and both 3C and Memento were included in the latest AD-GWAS (19). Summary statistics were re-generated for these GWAS meta anlayses with the overlapping cohorts removed. It was not possible to eliminate all sample overlap, as 3C-Dijon specifically had been included in a number of GWAS for cSVD MRI-markers and vascular risk factors. This consideration led us to choose 3C-Bordeaux-Montpellier as the training set as it was most important to limit sample overlap bias during the iPRS generation phase. We expect any remaining bias due to sample overlap with 3C-Dijon to be negligible.

eMethods 2. PRS Methods

In total, we generated 148 candidate PRS models per trait using different combinations of hyper-parameters for each method (70 C+T, 73 LDPRED-2, and 5 PRS-CS) in the 3C-Bordeaux-Montpellier training set.

Clumping and thresholding (C + T) is the standard approach used to generate PRS (20). In this method, SNPs are first clumped to account for LD. Specifically, clumping is an iterative process in which the most significant SNP in a specified genetic distance, *w*, is selected, and any variants correlated with this index variant beyond a defined r2 threshold are removed. Then a p-value threshold is applied to shrink effect sizes for SNPs with a p-value over the threshold to zero. This results in three hyper-parameters to select when applying C+T – the correlation threshold, r^2, and a p-value threshold (p_t). Typically, in PRS studies the r2 (0.1 or 0.2) are pre-set, and various p_t values are tested ranging from 1 to 5x10-8 (genome-wide significance level) (21). For candidate sPRS, we tune r^2 and p_t using a set of 70 different parameterizations:

- r^2 = 0.01, 0.1, 0.2, 0.5, 0.8
- p_t = $5x10^{-8}$, $1x10^{-6}$, $1x10^{-5}$, $1x10^{-4}$, 0.001, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.8, 1

The C+T approach was implemented using *PRSice-2* software (22).

C+T has noted limitations, namely that information is lost in the clumping process where potentially causal SNPs are discarded, and effect sizes set to zero based on a strict cut-off resulting in limited prediction accuracy (23). Ideally, PRS would be able to jointly model all SNPs across the genome to make full use of available information (24). Two Bayesian approaches to PRS, ***LDPred2* and *PRS-CS* software** attempt to directly model LD and limit loss of information (24,25).

LDpred2 is an improvement over the original LDpred method for deriving polygenic risk scores based only on an LD matrix (23,25). Essentially, the original LDpred method computes a LD-matrix that infers posterior mean effect sizes of each SNP by conditioning on a genetic architecture prior and LD information from an external reference panel[6]. We used the UK Biobank LD Reference panel, and conducted quality control, as recommended by the authors (26). LDpred2 requires the estimation of two parameters to compute the LD-matrix; (ρ) the fraction of causal SNPs (ρ), and SNP heritability (h^2_{SNP}) estimated from the GWAS summary statistics with LD score regression (8). LDpred2 additionally offers a 'sparse' option that learns which SNP-effects are truly zero, and an 'auto' option to automatically tune these parameters which does not require validation data and performs almost as well as a grid-model that takes considerably more computational effort. LDpred2 is implemented using the *bigsnpr* R package (27), and analysis was conducted closely following a tutorial provided by the authors (Polygenic scores and inference using LDpred2 bigsnpr (privefl.github.io)). We implemented the 'auto' model as well as the following combinations of tuning parameters (73 in total):

- ρ = $1x10^{-5}$, $2.8x10^{-5}$, $8.1x10^{-5}$, $2.3x10^{-4}$, $6.6x10^{-4}$, 0.0019, 0.0053, 0.015, 0.043, 0.12, 0.35, 1

- $h^2_{SNP} = h^2_{est}$, h^2_{est}*0.7, h^2_{est}*1.4
- Sparsity (Yes or No)

PRS-cs is a similar approach distinguished by utilizing a continuous shrinkage prior on SNP effect sizes (24). The use of a continuous shrinkage prior on SNP effect sizes is meant to be adaptive to the strength of its association signal in the GWAS and can accommodate for diverse underlying genetic architectures. The main difference with LDpred comes from jointly updating the effect sizes in each LD block in multivariate fashion versus updating effect sizes for each SNP separately and sequentially – leading to substantial computational improvements and better modeling of local LD patterns. PRS-cs requires tuning of the global shrinkage parameter (ϕ) that represents prior knowledge about the sparseness of the underlying genetic architecture. The software also offers an 'automatic' tuning option. PRS-cs is implemented using the 1000 Genomes European sample as an external LD reference (5). We implemented the following ϕ values as well as the 'auto' option:

- $\Phi = 1x10^{-6}, 1x10^{-4}, 0.01, 1$

eMethods 3. Construction of iPRS-DEM Model

We included all candidate PRS models for each trait in separate Fine-Gray regression models for all-cause dementia in the training set, adjusting for age, sex, study center, *APOE*-ε4 and ε2 dosages and 10 PCs. The formula for a sPRS for a given trait can be given as:

$$sPRS_n \propto \sum_{j=1}^{m} \alpha_j X_{ij}$$

where for trait *n*, *m* is the total number of SNPs in the sPRS, α_j is the effect size for the *j-th* SNP derived from the PRS model (some parameter combination from C+T, LDPRED2 or PRS-CS), and X_{i_j} is the number of risk alleles for *i*-th individual's *j*-th SNP. For each trait we selected the PRS with the lowest integrated Brier score (a common metric used to measure prediction error of survival models) across 5, 7.5 and 10 years (28,29). The Brier score is a measure of overall performance, including prediction and calibration, calculated by the average mean squared error at a given time point with lower scores indicating better performance (28). The integrated Brier score represents global average of the Brier score calculated at the aforementioned time points during follow-up. Brier scores were calculated with the .632+ bootstrapping method (1,000 replications) in the *riskRegression* R package (30,31). Traits were removed if no PRS model outperformed a benchmark model that included only *APOE*-ε4 and ε2 dosages (in addition to age, sex, PCs and study center). Each of the remaining single-trait PRS models were standardized (zero mean, unit standard deviation).

Next, to prevent overfitting and address correlation between single-trait PRS, we performed regularization by adapting elastic-net regression (32) to Fine-Gray. Elastic-net regression makes use of both LASSO and Ridge penalties to shrink coefficients and perform sparse selection (shrink some coefficients to zero). We adapted cox elastic-net regression to the Fine-Gray setting by first generating a special dataset using the *finegray* function from the *survival* R package (33), and performing a weighted cox regression using the functionalities of the *glmnet* R package (34). A range of Fine-Gray models (including all selected PRS for each trait and adjusted for age, sex, study centre, 10PCs, and *APOE*-ε4 and ε2 dosages) with different regularization parameters were evaluated using 25 repeats of 10-fold cross-validation. The best elastic-net model selected based on the lowest average cross-validated deviance (α=0.0787; λ=0.001163819). From this elastic-net model, we extracted the coefficients for each sPRS (γ_n), and with the empirical standard deviation (σ_n) derived training weights $(\frac{\gamma_1}{\sigma_1}, \ldots, \frac{\gamma_n}{\sigma_n})$ that were used to construct the final iPRS model:

$$iPRS_i \propto \sum_{j=1}^{m} X_{i_j}\left(\frac{\gamma_1}{\sigma_1}\alpha_{j_1} + \cdots + \frac{\gamma_n}{\sigma_n}\alpha_{j_n}\right)$$

where *m* is the total number of SNPs in the iPRS, $\sigma_1, \ldots, \sigma_n$ are the empirical standard deviations of each of the *n-th* sPRS in the training set, $\alpha_{j_1}, \ldots, \alpha_{j_n}$ are the effect size for the j-th SNP in each PRS model respectively, and X_{i_j} is the number of risk alleles for *i*-th individual's *j*-th SNP. A SNP's effect size α_{jn} was equal to zero in a given PRS if the SNP was not present. The final weight for each SNP in the iPRS-DEM was derived by summing across the re-weighted SNP effect sizes.

eMethods 4. Clinical risk factor definitions in 3C-Dijon

We aimed to compare iPRS-DEM against established, clinical risk factors for dementia. We assessed the following risk factors in 3C:

Risk Factor	Definition
Low Education	Did not complete beyond primary education
Smoking	Current smoker
Obesity	Body Mass Index ≥ 30 kg/m²
Hypertension	Blood pressure ≥ 140/90 mmHg or antihypertensive treatment
Diabetes	Fasting glycemia ≥ 7 mmol/L or taking diabetes medication
History of CVD	History of: myocardial infarction, hospitalized stroke, coronary surgery/angioplasty, history of leg artery surgery if arteritis of the lower limbs
Dyslipidemia	Fasting blood cholesterol ≥ 6.2 mmol/L or taking lipid-lowering medication

Those showing nominal association (P<0.05)with increased dementia risk were used in comparative prediction models. We constructed the following Fine-Gray models:

- Reference: age, sex, 10 genetic principal components (included in all subsequent models)
- APOE: Reference + *APOE*-ε4 and ε2 dosages
- APOE + AD-PRS/iPRS-DEM: Reference + *APOE*-ε4 and ε2 dosages + iPRS-DEM OR AD-PRS
- Individual Non-Genetic Risk Factor: Reference + clinical risk factor (e.g. Diabetes)
- Non-Genetic Factors: Reference + all non-genetic clinical risk factors
- APOE & Non-Genetic Factors: Reference + *APOE*-ε4 and ε2 dosages + all non-genetic clinical risk factors
- Full Model (iPRS-DEM): Reference + *APOE*-ε4 and ε2 dosages + all non-genetic clinical risk factors + iPRS-DEM
- Full-Model (AD-PRS): Reference + *APOE*-ε4 and ε2 dosages + all non-genetic clinical risk factors + AD-PRS

eMethods 5. Fine-Gray Regression Models and Interpretation

We employed Fine-Gray regression (FGR) to take into account competing risk of death (35). Fine-Gray models are widely used to model associations with cumulative incidence, or probability over a specified time period (i.e. risk). This approach is recommended for prediction as it overcomes limitations of Cox proportional hazard models to estimate the effect of covariates on the cumulative incidence function which describes the incidence of

the occurrence of an event while taking competing risk into account (36). However, Fine-Gray regression coefficients, corresponding to subdistribution hazard ratios, have been found to be widely misinterpreted across studies (37). Importantly, in Fine-Gray models, the 'subdistriubtion' hazard function is defined as the instantaneous rate of occurrence of the event of interest (dementia) among participants who have not currently experienced the event. In other words, we consider participants who have experienced the competing event (death) as part of the risk set, even if they contribute 'immortal' time (i.e. experienced the competing event). This is necessary to maintain an explicit one-to-one link to the cumulative incidence function with the implication that if a covariate increases the subdistribution hazard, it will also increase the incidence of the event. However, while the direction of association can be described, the magnitude of this effect cannot be directly quantified. One can only make statements on relative magnitude of effects between different covariates on the incidence of the same outcome within the same study. Simulations have shown that subdistribution hazard ratios may approximate the effect on the risk of the event when the probability of the event is relatively infrequent (<20%) (37).

As such, we caution against making comparisons across 3C-Dijon and Memento in terms of the strength of association of iPRS-DEM. Further, any interpretation of magnitude of effect on risk of dementia can only be approximated since the cumulative incidence of dementia in each study reaches only 9.8% over 10-years follow-up in 3C-Dijon, and 14% over 5-years follow-up in Memento.

eMethods 6. Age at onset analysis in 3C-Dijon and Memento

In order to assess age at onset in longitudinal cohorts, we performed two separate analyses using age at censoring, death, or dementia before/after 80 years of age. For the <80 analysis, participants with a baseline age >80 were removed, and participants were censored based on their status at age 80 (i.e. if an individual was diagnosed with dementia after this date, they instead censored as non-demented in this analysis). Conversely, for the >80 analysis, we considered 'baseline' as age 80 and included only participants who were dementia-free (or alive) at age 80 and had follow-up data after this age. In both analyses the association of iPRS-DEM with cumulative incidence of dementia was adjusted for baseline covariates (age at baseline, sex, 10 PCs) and APOE-ε4 and ε2 dosage.

- Resulting number of participants and cases in the age at onset analyses can be found in eTable 10 (3C-Dijon) and eTable 15 (Memento)

eMethods 7. Defining optimal cohort-specific cutoffs to dichotomize iPRS-DEM

We sought to determine the optimal iPRS-DEM cutoff in *APOE*-ε4 stratum separately since underlying dementia pathology may differ substantially in carriers vs. non-carriers (38). Indeed, iPRS-DEM could be capturing genetic risk related to different pathological mechanisms dependent on *APOE*-ε4. The objective of this analysis was not only to identify genetic risk groups defined by iPRS-DEM and *APOE*-ε4 status, but to additionally gain insight into if risk is distributed differentially across iPRS-DEM according to *APOE*-ε4 strata.

iPRS-DEM cutoffs were identified within each *APOE*-ε4 stratum with a validated algorithm to dichtomize continuous variables in multivariable survival models (39). Briefly, we used the *findcutCox* R function provided by Chang et al. that utilizes an efficient genetic algorithm to optimize the likelihood ratio test statistic in cox proporitional hazard models, and estimate the location of cut points. We constrained the cut point between the 50th and 95th percentiles to maintain reasonable sample sizes in genetic risk groups, and adjusted for age at baseline, sex, and 10 PCs.

In Memento we applied the same procedure. We decided to use cohort-specific cutoffs since optimal dichotomization of iPRS-DEM may not be generalizable between a healthier community-based and high-risk memory clinic population.

eMethods 8. Cumulative Incidence of all-cause dementia in genetic risk strata in 3C-Dijon

We conducted a non-parametric analysis of cumulative incidence of dementia across different ages in 3C-Dijon. Here, we estimated cumulative incidence of dementia, accounting for death as a competing event, within each of the four defined genetic risk strata, while accounting for left truncation. Briefly, accounting for left truncation (that participants are not observable prior to baseline) avoided potential under-estimation of cumulative incidence of dementia by considering individuals to enter the risk set at age of baseline assessment. We chose age as a time scale for this analysis as, in a population-based cohort such as 3C, it is useful to understand possible differences in cumulative incidence at different ages rather than from baseline assessment. Further, in the 4 genetic risk groups defined by APOE-ε4 and iPRS-DEM, differences in cumulative incidence between were compared at 75, 80, 85, and 90 years old using an approximate standard normal distribution under the null hypothesis that estimated incidences are equivalent between groups (40).

eMethods 9. Risk prediction models in 3C-Dijon

Clinical risk factors showing nominal association (P<0.05)with increased dementia risk were used in comparative prediction models. We constructed the following Fine-Gray models:

- Reference: age, sex, 10 genetic principal components (included in all subsequent models)
- APOE: Reference + *APOE*-ε4 and ε2 dosages
- APOE + AD-PRS/iPRS-DEM: Reference + *APOE*-ε4 and ε2 dosages + iPRS-DEM OR AD-PRS
- Individual Non-Genetic Risk Factor: Reference + clinical risk factor (e.g. Diabetes)
- Non-Genetic Factors: Reference + all non-genetic clinical risk factors
- APOE & Non-Genetic Factors: Reference + *APOE*-ε4 and ε2 dosages + all non-genetic clinical risk factors
- Full Model (iPRS-DEM): Reference + *APOE*-ε4 and ε2 dosages + all non-genetic clinical risk factors + iPRS-DEM
- Full-Model (AD-PRS): Reference + *APOE*-ε4 and ε2 dosages + all non-genetic clinical risk factors + AD-PRS

eMethods 10. Clinical risk factor definitions and risk prediction models in Memento

We used a linear term for age at baseline since the age-squared term had worse fit according to AIC. This is likely due to Memento consisting of a larger age range with more younger individuals, and a higher burden of dementia risk than 3C leading to younger onset. We investigated the following risk factors in Memento:

Risk Factor	Definition
Low Education	Did not complete beyond primary education
Family History of Dementia	2nd degree relative diagnosed with dementia: grandparent, mother, father, child, brother, sister, uncle, aunt or first cousin
Hypertension	Blood pressure ≥ 140/90 mmHg or antihypertensive treatment
Smoking	Current Smoker
Diabetes	diagnosed diabetes at baseline assessment
Dyslipidemia	dyslipidemia treated at inclusion

History of Cardiovascular Disease	anyone with recorded coronary artery disease, stroke, myocardial infarction, or cardiac valve issues in their medical history
Obesity	Body Mass Index ≥ 30 kg/m²

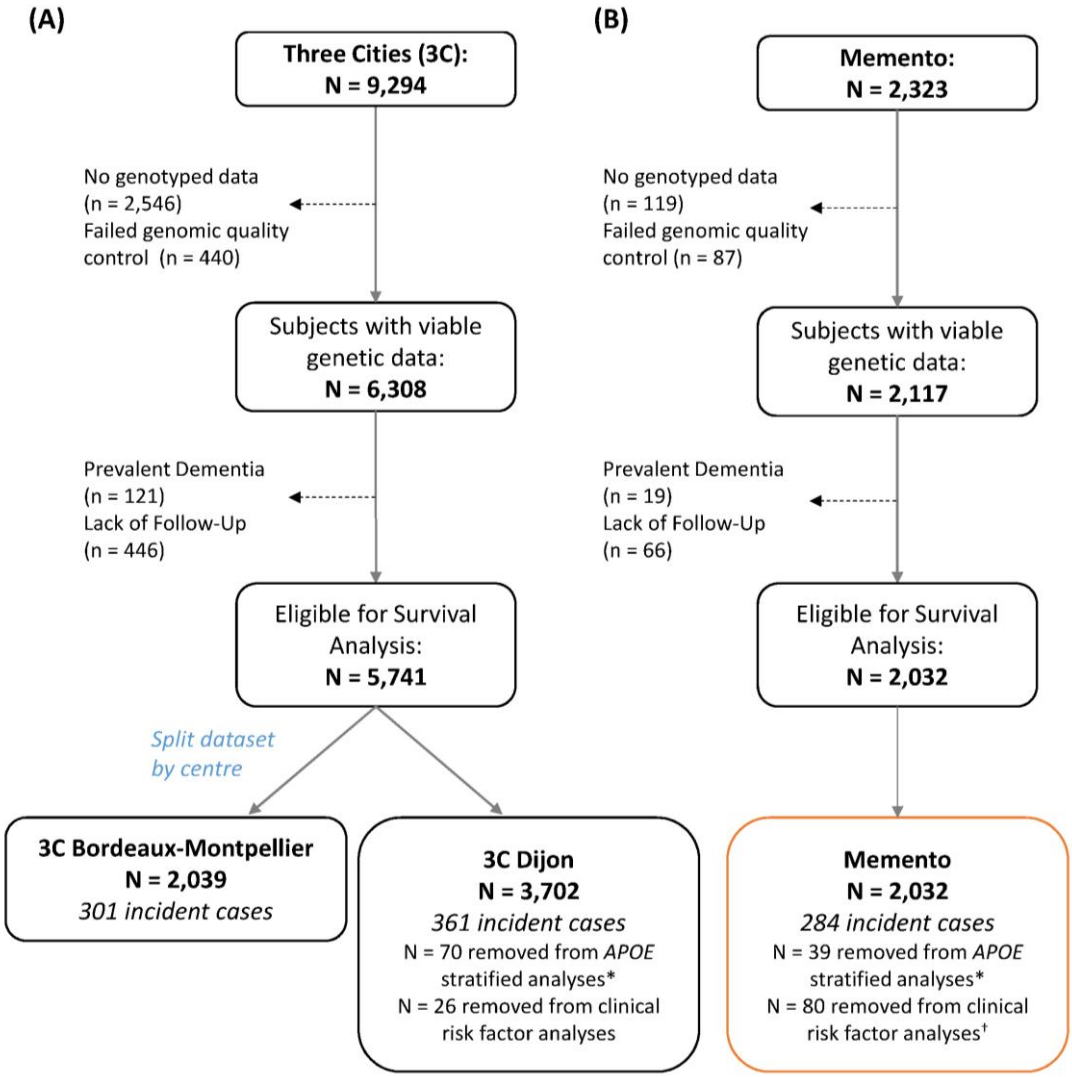
We again limited prediction models to include risk factors nominally associated with increased dementia risk ($P<0.05$). Prediction performance in Memento was evaluated at 5-years follow-up.

eMethods 11. Cumulative incidence estimation in Memento

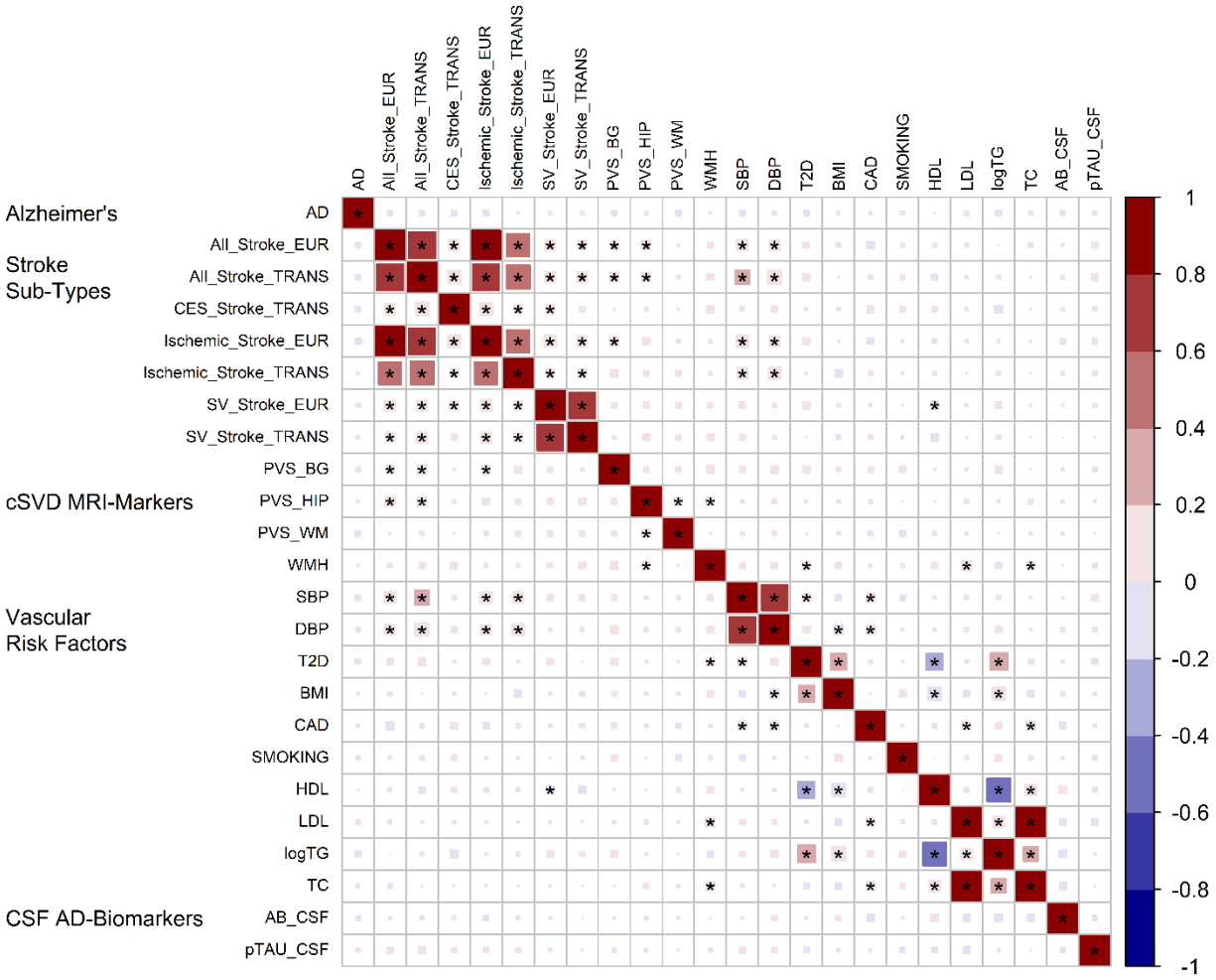
We used time since baseline assessment since progression from cognitive complaints to dementia is of interest in this clinical population and, statistically, overly small risk sets at younger ages lead to instability in cumulative incidence estimation (41).The *adjustedCurves* R package was used to construct cumulative incidence curves adjusted for baseline covariates (age, sex, and 10 PCs) using the inverse weighted transition probability method and to test differences between risk estimates at each year since study entry (42,43). Briefly, the inverse weighted transition probability method estimates propensity scores for each individual belonging to each group (genetic risk strata) based on the set of adjustment variables, and the inverse of this probability is calculated and used as weights to adjust the analysis. Here, we used a multinomial model for belonging to any of the 4 genetic risk strata to estimate propensity scores. We tested differences between estimated cumulative incidence of genetic risk groups at 1, 2, 3 , 4, and 5-years follow-up.

Supplementary Figures

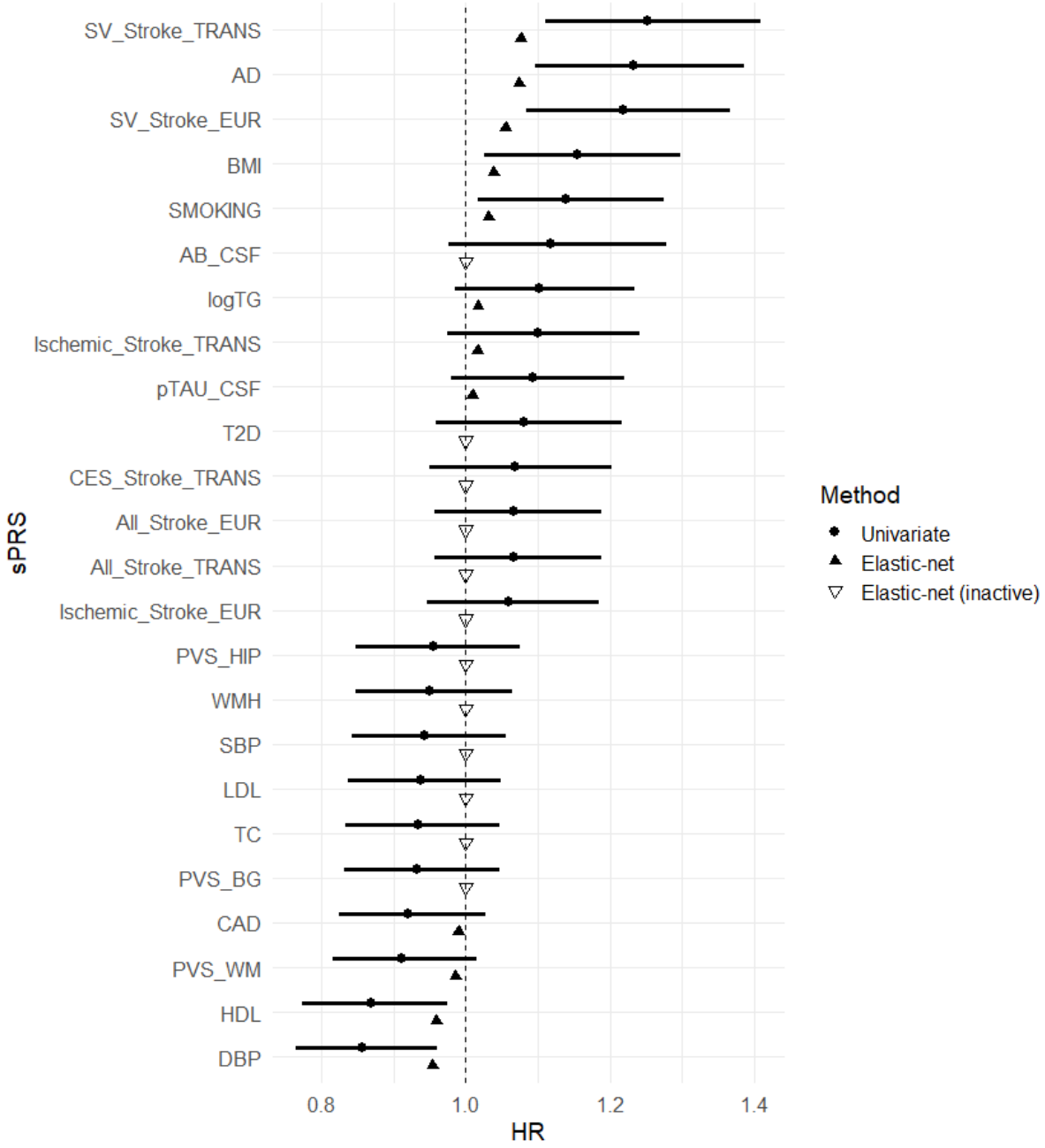
eFigure 1. Participant exclusion flowchart for (A) Three Cities and (B) Memento. **The full 3C-Dijon and Memento analytical sample were used for main association analyses. We performed complete cases analysis for all analyses including clinical risk factors. *In *APOE-ε4* stratified analyses we removed APOE-*ε2/ε4* individuals.†In Memento only low education and family history of dementia were used in prediction models since other risk factors was not associated with increased dementia risk, only 11 individuals were removed from predictive models.**



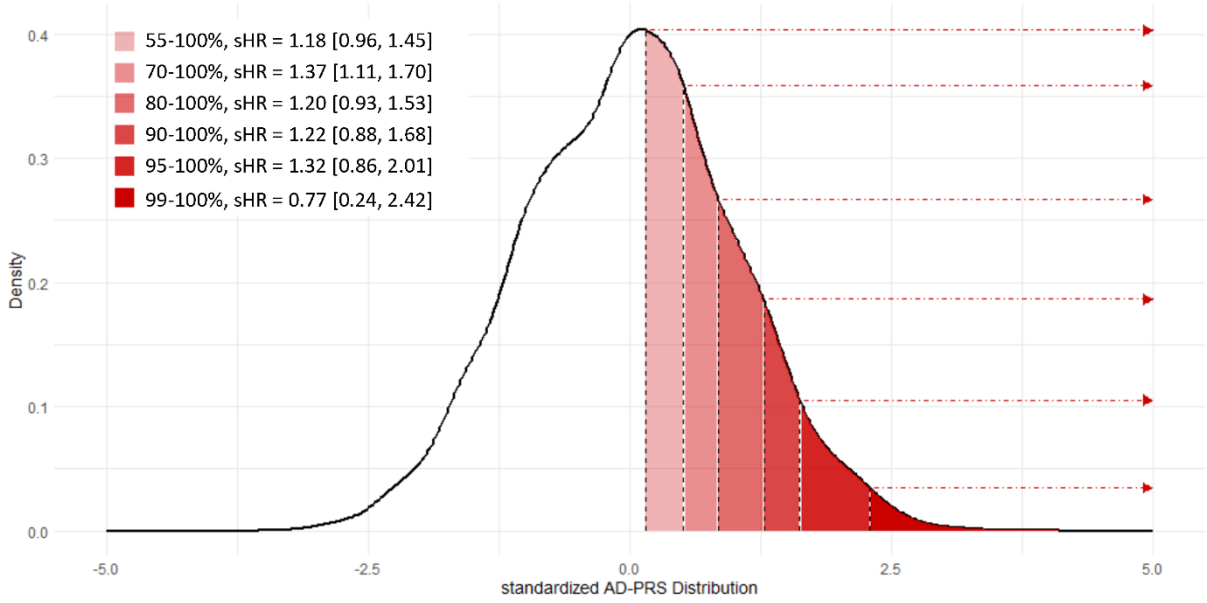
eFigure 2. Correlations between all single-trait PRS included in iPRS-DEM. *indicates nominally significant correlations (p<0.05).



eFigure 3. sPRS associations with cumulative incidence of dementia before and after Fine-Gray Elastic-net Regularization in training set.
All models are adjusted for age, sex, study centre, 10 PCs, and APOE-ε4 and APOE-ε2 dosages. Univariate indicates only one single-trait PRS was added to the model. Traits were determined 'inactive' if they had a training weight of zero. See eTable 3 for precise training weights.



eFigure 4. Association of AD-PRS with cumulative incidence of dementia in 3C-Dijon across percentile groups. Models are adjusted for age, sex, 10 PCs, and APOE-ε4 and APOE-ε2 dosages. At each percentile (55th, 70th, etc.) the AD-PRS is dichotomized.



References

1. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. Nat Protoc. 2010 Sep;5(9):1564–73.

2. Marees AT, de Kluiver H, Stringer S, Vorspan F, Curis E, Marie-Claire C, et al. A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. Int J Methods Psychiatr Res. 2018;27(2):e1608.

3. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. Nat Genet. 2016 Oct;48(10):1284–7.

4. Pain O, Glanville KP, Hagenaars SP, Selzam S, Fürtjes AE, Gaspar HA, et al. Evaluation of polygenic prediction methodology within a reference-standardized framework. Plagnol V, editor. PLOS Genet. 2021 May 4;17(5):e1009021.

5. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, et al. A global reference for human genetic variation. Nature. 2015 Oct;526(7571):68–74.

6. All of Us Research Program: Genomic Research Data Quality Report. [Internet]. National Institutes of Health; [cited 2024 Feb 4]. Available from: https://support.researchallofus.org/hc/en-us/article_attachments/19370367115796

7. Bianchi DW, Brennan PF, Chiang MF, Criswell LA, D'Souza RN, Gibbons GH, et al. The All of Us Research Program is an opportunity to enhance the diversity of US biomedical research. Nat Med. 2024 Feb;30(2):330–3.

8. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet. 2015 Nov;47(11):1236–41.

9. Choi SW, Mak TSH, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. Nat Protoc. 2020 Sep;15(9):2759–72.

10. Purcell S, Chang CC. PLINK [1.90] [Internet]. Available from: www.cog-genomics.org/plink/1.9/

11. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 2015 Dec;4(1):7.

12. Leonenko G, Baker E, Stevenson-Hoare J, Sierksma A, Fiers M, Williams J, et al. Identifying individuals with high risk of Alzheimer's disease using polygenic risk scores. Nat Commun. 2021 Jul 23;12(1):4506.

13. Mishra A, Malik R, Hachiya T, Jürgenson T, Namba S, Posner DC, et al. Stroke genetics informs drug discovery and risk prediction across ancestries. Nature. 2022 Nov;611(7934):115–23.

14. Chauhan G, Adams HHH, Satizabal CL, Bis JC, Teumer A, Sargurupremraj M, et al. Genetic and lifestyle risk factors for MRI-defined brain infarcts in a population-based setting. Neurology. 2019 Jan 29;92(5):e486–503.

15. Knol MJ, Lu D, Traylor M, Adams HHH, Romero JRJ, Smith AV, et al. Association of common genetic variants with brain microbleeds. Neurology. 2020 Dec 15;95(24):e3331–43.

16. Brouwer RM, Klein M, Grasby KL, Schnack HG, Jahanshad N, Teeuw J, et al. Genetic variants associated with longitudinal changes in brain structure across the lifespan. Nat Neurosci. 2022 Apr;25(4):421–32.

17. Choi SW, Mak TSH, Hoggart CJ, O'Reilly PF. EraSOR: a software tool to eliminate inflation caused by sample overlap in polygenic score analyses. GigaScience. 2023 Jan 1;12:giad043.

18. Wray NR, Yang J, Hayes BJ, Price AL, Goddard ME, Visscher PM. Pitfalls of predicting complex traits from SNPs. Nat Rev Genet. 2013 Jul;14(7):507–15.

19. Bellenguez C, Küçükali F, Jansen IE, Kleineidam L, Moreno-Grau S, Amin N, et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. Nat Genet. 2022 Apr 1;54(4):412–36.

20. Privé F, Vilhjálmsson BJ, Aschard H, Blum MGB. Making the Most of Clumping and Thresholding for Polygenic Scores. Am J Hum Genet. 2019 Dec 5;105(6):1213–21.

21. Wray NR, Lee SH, Mehta D, Vinkhuyzen AAE, Dudbridge F, Middeldorp CM. Research Review: Polygenic methods and their application to psychiatric traits. J Child Psychol Psychiatry. 2014;55(10):1068–87.

22. Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data. GigaScience. 2019 Jul 1;8(7):giz082.

23. Vilhjálmsson BJ, Yang J, Finucane HK, Gusev A, Lindström S, Ripke S, et al. Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. Am J Hum Genet. 2015 Oct;97(4):576–92.

24. Ge T, Chen CY, Ni Y, Feng YCA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. Nat Commun. 2019 Apr 16;10(1):1776.

25. Privé F, Arbel J, Vilhjálmsson BJ. LDpred2: better, faster, stronger. Bioinformatics. 2021 Apr 1;36(22–23):5424–31.

26. Privé F, Arbel J, Aschard H, Vilhjálmsson BJ. Identifying and correcting for misspecifications in GWAS summary statistics and polygenic scores. Hum Genet Genomics Adv. 2022 Oct 13;3(4):100136.

27. Privé F, Aschard H, Ziyatdinov A, Blum MGB. Efficient analysis of large-scale genome-wide data with two R packages: bigstatsr and bigsnpr. Bioinformatics. 2018 Aug 15;34(16):2781–7.

28. Gerds TA, Schumacher M. Consistent Estimation of the Expected Brier Score in General Survival Models with Right-Censored Event Times. Biom J. 2006;48(6):1029–40.

29. Park SY, Park JE, Kim H, Park SH. Review of Statistical Methods for Evaluating the Performance of Survival or Other Time-to-Event Prediction Models (from Conventional to Deep Learning Approaches). Korean J Radiol. 2021 Oct;22(10):1697–707.

30. Gerds TA, Kattan MW. Medical Risk Prediction Models: With Ties to Machine Learning (1st ed.). Chapman and Hall/CRC; 2021.

31. Gerds TA, Ohlendorff JS, Ozenne B. riskRegression: Risk Regression Models and Prediction Scores for Survival Analysis with Competing Risks [Internet]. 2023. Available from: https://CRAN.R-project.org/package=riskRegression

32. Zou H, Hastie T. Regularization and Variable Selection Via the Elastic Net. J R Stat Soc Ser B Stat Methodol. 2005 Apr 1;67(2):301–20.

33. Therneau T. A package for survival analysis in R [Internet]. 2024. Available from: https://CRAN.R-project.org/package=survival

34. Simon N, Friedman JH, Hastie T, Tibshirani R. Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. J Stat Softw. 2011 Mar 9;39:1–13.

35. Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. J Am Stat Assoc. 1999 Jun 1;94(446):496–509.

36. Austin PC, Lee DS, Fine JP. Introduction to the Analysis of Survival Data in the Presence of Competing Risks. Circulation. 2016 Feb 9;133(6):601–9.

37. Austin PC, Fine JP. Practical recommendations for reporting Fine-Gray model analyses for competing risk data. Stat Med. 2017;36(27):4391–400.

38. McCorkindale AN, Mundell HD, Guennewig B, Sutherland GT. Vascular Dysfunction Is Central to Alzheimer's Disease Pathogenesis in APOE e4 Carriers. Int J Mol Sci. 2022 Jun 26;23(13):7106.

39. Chang C, Hsieh MK, Chiang AJ, Tsai YH, Liu CC, Chen J. Methods for estimating the optimal number and location of cut points in multivariate survival analysis: a statistical solution to the controversial effect of BMI. Comput Stat. 2019 Dec 1;34(4):1649–74.

40. Meister R, Schaefer C. Statistical methods for estimating the probability of spontaneous abortion in observational studies—Analyzing pregnancies exposed to coumarin derivatives. Reprod Toxicol. 2008 Sep 1;26(1):31–5.

41. Stegherr R, Allignol A, Meister R, Schaefer C, Beyersmann J. Estimating cumulative incidence functions in competing risks data with dependent left-truncation. Stat Med. 2020;39(4):481–93.

42. Denz R, Klaaßen-Mielke R, Timmesfeld N. A comparison of different methods to adjust survival curves for confounders. Stat Med. 2023;42(10):1461–79.

43. Neumann A, Billionnet C. Covariate adjustment of cumulative incidence functions for competing risks data using inverse probability of treatment weighting. Comput Methods Programs Biomed. 2016 Jun;129:63–70.

Data availability

The iPRS-DEM will be made publicly available on the PGS catalog (https://www.pgscatalog.org/) upon publication. We used publicly available data for analyses described in this manuscript, including data from GWAS catalog (https://www.ebi.ac.uk/gwas/, study code: GCST90244151, GCST006630, GCST009004, GCST007474, GCST90129599, GCST90129600). GWAS summary statistics from the Global Lipids Genetics Consortium for lipid measurements were publicly available (http://csg.sph.umich.edu/willer/public/glgc-lipids2021/). GWAS for type-2 diabetes from the DIAGRAM, DIAMANTE, and T2DGGI consortia are available at http://diagram-consortium.org/downloads.html. White matter hyperintensities GWAS summary statsitics are available at dbGaP (https://www.ncbi.nlm.nih.gov/gap/), under the accession number: phs002227.v1.p1. Cerebral Microbleeds GWAS summary statistics are available at under the accession number phs000930.v7.p1. All other GWAS summary statistics were provided by original authors upon request.

Code availability

We used publically available software and tools to conduct analyses. PLINK1.90 and PLINK2.0 (available at https://www.cog-genomics.org/plink/; https://www.cog-

genomics.org/plink/2.0/) was used to perform quality control procedures of genotyped data. Scripts for further quality control and harmonization of GWAS summary statistics and the 1000G reference panel were adapted from the GenoPred pipeline (https://opain.github.io/GenoPred/). To generate polygenic risk score models we used publicly available tools: PRSice-2 (https://choishingwan.github.io/PRSice/), LDpred-2 from the bigsnpr R package (https://github.com/privefl/bigsnpr/), and PRS-CS (https://github.com/getian107/PRScs). R software can be downloaded from (https://www.r-project.org/). We used the following R packages for statistical analyses: MungeSumstats (https://github.com/neurogenomics/MungeSumstats), riskRegression (https://github.com/tagteam/riskRegression), glmnet (https://glmnet.stanford.edu/articles/glmnet.html), etm (https://github.com/cran/etm), adjustedCurves (https://github.com/RobinDenz1/adjustedCurves), and survival (https://github.com/therneau/survival). Figures were created using ggplot2 R package (https://ggplot2.tidyverse.org/).

Participating consortia

The GIGASTROKE Consortium

Aniket Mishra[1], Rainer Malik[2], Tsuyoshi Hachiya[3], Tuuli Jürgenson[4,5], Shinichi Namba[6], Daniel C. Posner[7], Frederick K. Kamanu[8,9], Masaru Koido[10,11], Quentin Le Grand[1], Mingyang Shi[11], Yunye He[11], Marios K. Georgakis[2,12,13], Ilana Caro[1], Kristi Krebs[4], Yi-Ching Liaw[14,15], Felix C. Vaura[16,17], Kuang Lin[18], Bendik Slagsvold Winsvold[19,20,21], Vinodh Srinivasasainagendra[22], Livia Parodi[12,13], Hee-Joon Bae[23], Ganesh Chauhan[24], Michael R. Chong[25,26], Liisa Tomppo[27], Rufus Akinyemi[28,29], Gennady V. Roshchupkin[30,31], Naomi Habib[32], Yon Ho Jee[33], Jesper Qvist Thomassen[34], Vida Abedi[35,36], Jara Cárcel-Márquez[37,38], Marianne Nygaard[39,40], Hampton L. Leonard[41,42,43], Chaojie Yang[44,45], Ekaterina Yonova-Doing[46,47], Maria J. Knol[30], Adam J. Lewis[48], Renae L. Judy[49], Tetsuro Ago[50], Philippe Amouyel[51,52,53], Nicole D. Armstrong[54], Mark K. Bakker[55], Traci M. Bartz[56,57], David A. Bennett[58], Joshua C. Bis[56], Constance Bordes[1], Sigrid Børte[20,59,60], Anael Cain[32], Paul M. Ridker[61,62], Kelly Cho[7], Zhengming Chen[18,63], Carlos Cruchaga[64,65], John W. Cole[66,67], Phil L. de Jager[13,68], Rafael de Cid[69], Matthias Endres[70,71,72,73], Leslie E. Ferreira[74], Mirjam I. Geerlings[75], Natalie C. Gasca[57], Vilmundur Gudnason[76,77], Jun Hata[78], Jing He[48], Alicia K. Heath[79], Yuk-Lam Ho[7], Aki S. Havulinna[80,81], Jemma C. Hopewell[82], Hyacinth I. Hyacinth[83], Michael Inouye[46,84,85,86,87], Mina A. Jacob[88], Christina E. Jeon[89], Christina Jern[90,91], Masahiro Kamouchi[92], Keith L. Keene[93], Takanari Kitazono[50], Steven J. Kittner[67,94], Takahiro Konuma[6], Amit Kumar[24], Paul Lacaze[95], Lenore J. Launer[96], Keon-Joo Lee[97], Kaido Lepik[4,98,99,100], Jiang Li[35], Liming Li[101], Ani Manichaikul[44], Hugh S. Markus[102], Nicholas A. Marston[8,9], Thomas Meitinger[103,104], Braxton D. Mitchell[105,106], Felipe A. Montellano[107,108], Takayuki Morisaki[10], Thomas H. Mosley[109], Mike A. Nalls[41,42,43], Børge G. Nordestgaard[110,111], Martin J. O'Donnell[112], Yukinori Okada[6,113,114,115,116,117], N. Charlotte Onland-Moret[75], Bruce Ovbiagele[118], Annette Peters[119,120,121], Bruce M. Psaty[56,122,123], Stephen S. Rich[44], Jonathan Rosand[12,13,124], Marc S. Sabatine[8,9], Ralph L. Sacco[125,126], Danish Saleheen[127], Else Charlotte Sandset[128,129], Veikko Salomaa[80], Muralidharan Sargurupremraj[130], Makoto Sasaki[3], Claudia L. Satizabal[130,131], Carsten O. Schmidt[132], Atsushi Shimizu[3], Nicholas L. Smith[122,133,134], Kelly L. Sloane[135], Yoichi Sutoh[3], Yan V. Sun[136,137], Kozo Tanno[3], Steffen Tiedt[2], Turgut Tatlisumak[138], Nuria P. Torres-Aguila[37], Hemant K. Tiwari[22], David-Alexandre Trégouët[1], Stella Trompet[139,140], Anil Man Tuladhar[88], Anne Tybjærg-Hansen[34,111], Marion van Vugt[141], Riina Vibo[142], Shefali S. Verma[143], Kerri L. Wiggins[56], Patrik Wennberg[144], Daniel Woo[83], Peter W. F. Wilson[136,145], Huichun Xu[105], Qiong Yang[131,146], Kyungheon Yoon[147], The COMPASS Consortium*, The INVENT Consortium*, The Dutch Parelsnoer Initiative (PSI) Cerebrovascular Disease Study Group*, The Estonian Biobank*, The PRECISE4Q Consortium*, The FinnGen Consortium*, The NINDS Stroke Genetics Network (SiGN)*, The MEGASTROKE Consortium*, The SIREN Consortium*, The China Kadoorie Biobank Collaborative Group*, The VA Million Veteran Program*, The International Stroke Genetics Consortium (ISGC)*, The Biobank Japan*, The CHARGE Consortium*, The GIGASTROKE Consortium*, Iona Y. Millwood[18,63], Christian Gieger[148], Toshiharu Ninomiya[78], Hans J. Grabe[149,150], J. Wouter Jukema[140,151,152], Ina L. Rissanen[75], Daniel Strbian[27], Young Jin Kim[147], Pei-Hsin Chen[15], Ernst Mayerhofer[12,13], Joanna M. M. Howson[46,47], Marguerite R. Irvin[54], Hieab Adams[153,154], Sylvia Wassertheil-Smoller[155], Kaare Christensen[39,40,156], Mohammad A. Ikram[30], Tatjana Rundek[125,126], Bradford B. Worrall[157,158], G. Mark Lathrop[159], Moeen Riaz[95],

Eleanor M. Simonsick[160], Janika Kõrv[142], Paulo H. C. França[74], Ramin Zand[161,162], Kameshwar Prasad[24], Ruth Frikke-Schmidt[34,111], Frank-Erik de Leeuw[88], Thomas Liman[71,75,163], Karl Georg Haeusler[108], Ynte M. Ruigrok[55], Peter Ulrich Heuschmann[107,164,165], W. T. Longstreth[122,166], Keum Ji Jung[18,167], Lisa Bastarache[48], Guillaume Paré[25,26,168,169], Scott M. Damrauer[170,171], Daniel I. Chasman[61,62], Jerome I. Rotter[172], Christopher D. Anderson[12,13,124,173], John-Anker Zwart[19,20,59], Teemu J. Niiranen[16,17,174], Myriam Fornage[175,176], Yung-Po Liaw[15,177], Sudha Seshadri[130,131,178], Israel Fernández-Cadenas[37], Robin G. Walters[18,63], Christian T. Ruff[8,9,419], Mayowa O. Owolabi[28,179,419], Jennifer E. Huffman[7,419], Lili Milani[4,419], Yoichiro Kamatani[11,419], Martin Dichgans[2,180,181,419] & Stephanie Debette[1,182,419]

[1]Bordeaux Population Health Research Center, University of Bordeaux, Inserm, UMR [1219], Bordeaux, France. [2]Institute for Stroke and Dementia Research (ISD), University Hospital, LMU Munich, Munich, Germany. [3]Iwate Tohoku Medical Megabank Organization, Iwate Medical University, Iwate, Japan. [4]Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia. [5]Institute of Mathematics and Statistics, University of Tartu, Tartu, Estonia. [6]Department of Statistical Genetics, Osaka University Graduate School of Medicine, Suita, Japan. [7]Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC), VA Boston Healthcare System, Boston, MA, USA. [8]TIMI Study Group, Boston, MA, USA. [9]Division of Cardiovascular Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [10]Division of Molecular Pathology, Institute of Medical Sciences, The University of Tokyo, Tokyo, Japan. [11]Laboratory of Complex Trait Genomics, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan. [12]Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA. [13]Program in Medical and Population Genetics, Broad Institute of Harvard and the Massachusetts Institute of Technology, Cambridge, MA, USA. [14]Laboratory of Clinical Genome Sequencing, Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan. [15]Department of Public Health and Institute of Public Health, Chung Shan Medical University, Taichung, Taiwan. [16]Department of Internal Medicine, University of Turku, Turku, Finland. [17]Department of Public Health and Welfare, Finnish Institute for Health and Welfare, Turku, Finland. [18]Nuffield Department of Population Health, University of Oxford, Oxford, UK. [19]Department of Research and Innovation, Division of Clinical Neuroscience, Oslo University Hospital, Oslo, Norway. [20]K. G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology (NTNU), Trondheim, Norway. [21]Department of Neurology, Oslo University Hospital, Oslo, Norway. [22]Department of Biostatistics, School of Public Health, University of Alabama at Birmingham, Birmingham, AL, USA. [23]Department of Neurology and Cerebrovascular Disease Center, Seoul National University Bundang Hospital, Seoul National University College of Medicine, Seongnam, Republic of Korea. [24]Rajendra Institute of Medical Sciences, Ranchi, India. [25]Thrombosis and Atherosclerosis Research Institute, David Braley Cardiac, Vascular and Stroke Research Institute, Hamilton, Ontario, Canada. [26]Department of Pathology and Molecular Medicine, Michael G. DeGroote School of Medicine, McMaster University, Hamilton, Ontario, Canada. [27]Department of Neurology, Helsinki University Hospital and University of Helsinki, Helsinki, Finland. [28]Center for Genomic and Precision Medicine, College of Medicine, University of Ibadan, Ibadan, Nigeria. [29]Neuroscience and Ageing Research Unit Institute for Advanced Medical Research and Training, College of Medicine, University of Ibadan, Ibadan, Nigeria. [30]Department of Epidemiology, Erasmus MC University Medical Center Rotterdam, Rotterdam, The Netherlands. [31]Department of Radiology and Nuclear Medicine, Erasmus MC University Medical Center Rotterdam, Rotterdam, The Netherlands. [32]The Edmond and Lily Safra Center for Brain Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel. [33]Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA, USA. [34]Department of Clinical Biochemistry, Copenhagen University Hospital—Rigshospitalet, Copenhagen, Denmark. [35]Department of Molecular and Functional Genomics, Weis Center for Research, Geisinger Health System, Danville, VA, USA. [36]Department of Public Health Sciences, College of Medicine, The Pennsylvania State University, State College, PA, USA. [37]Stroke Pharmacogenomics and Genetics Laboratory, Biomedical Research Institute Sant Pau (IIB Sant Pau), Barcelona, Spain. [38]Departament de Medicina, Universitat Autònoma de Barcelona, Barcelona, Spain. [39]The Danish Twin Registry, Department of Public Health, University of Southern Denmark, Odense, Denmark. [40]Department of Clinical Genetics, Odense University Hospital, Odense, Denmark. [41]Center for Alzheimer's and Related Dementias, National Institutes of Health, Bethesda, MD, USA. [42]Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA. [43]Data Tecnica International, Glen Echo, MD, USA. [44]Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA. [45]Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, VA, USA. [46]British Heart Foundation Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [47]Department of Genetics, Novo Nordisk Research Centre Oxford,

Oxford, UK. [48]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA. [49]Department of Surgery, University of Pennsylvania, Philadelphia, PA, USA. [50]Department of Medicine and Clinical Science, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan. [51]University of Lille, INSERM U[1167], RID-AGE, LabEx DISTALZ, Risk Factors and Molecular Determinants of Aging-Related Diseases, Lille, France. [52]CHU Lille, Public Health Department, Lille, France. [53]Institut Pasteur de Lille, Lille, France. [54]Department of Epidemiology, University of Alabama at Birmingham, Birmingham, AL, USA. [55]UMC Utrecht Brain Center, Department of Neurology and Neurosurgery, University Medical Center Utrecht, University Utrecht, Utrecht, The Netherlands. [56]Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA. [57]Department of Biostatistics, University of Washington, Seattle, WA, USA. [58]Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, USA. [59]Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway. [60]Research and Communication Unit for Musculoskeletal Health (FORMI), Department of Research and Innovation, Division of Clinical Neuroscience, Oslo University Hospital, Oslo, Norway. [61]Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA, USA. [62]Harvard Medical School, Boston, MA, USA. [63]MRC Population Health Research Unit, University of Oxford, Oxford, UK. [64]Department of Psychiatry, Washington University School of Medicine, Saint Louis, MO, USA. [65]NeuroGenomics and Informatics Center, Washington University School of Medicine, Saint Louis, MO, USA. [66]VA Maryland Health Care System, Baltimore, MD, USA. [67]Department of Neurology, University of Maryland School of Medicine, Baltimore, MD, USA. [68]Center for Translational and Computational Neuroimmunology, Department of Neurology, Columbia University Medical Center, New York, NY, USA. [69]GenomesForLifeGCAT Lab Group, Germans Trias i Pujol Research Institute (IGTP), Badalona, Spain. [70]Klinik und Hochschulambulanz für Neurologie, Charité—Universitätsmedizin Berlin, Berlin, Germany. [71]Center for Stroke Research Berlin, Berlin, Germany. [72]German Center for Neurodegenerative Diseases (DZNE), partner site Berlin, Berlin, Germany. [73] German Centre for Cardiovascular Research (DZHK), partner site Berlin, Berlin, Germany. [74]Post-Graduation Program on Health and Environment, Department of Medicine and Joinville Stroke Biobank, University of the Region of Joinville, Santa Catarina, Brazil. [75]Department of Epidemiology, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands. [76]Icelandic Heart Association, Kopavogur, Iceland. [77]Faculty of Medicine, University of Iceland, Reykjavik, Iceland. [78]Department of Epidemiology and Public Health, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan. [79]Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK. [80]Department of Public Health and Welfare, Finnish Institute for Health and Welfare, Helsinki, Finland. [81]Institute for Molecular Medicine Finland, FIMM-HiLIFE, Helsinki, Finland. [82]Clinical Trial Service and Epidemiological Studies Unit (CTSU), Nuffield Department of Population Health, University of Oxford, Oxford, UK. [83]Department of Neurology and Rehabilitation Medicine, University of Cincinnati College of Medicine, Cincinnati, OH, USA. [84]Cambridge Baker Systems Genomics Initiative, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [85]Cambridge Baker Systems Genomics Initiative, Baker Heart and Diabetes Institute, Melbourne, Victoria, Australia. [86]Health Data Research UK Cambridge, Wellcome Genome Campus and University of Cambridge, Cambridge, UK. [87]British Heart Foundation Centre of Research Excellence, University of Cambridge, Cambridge, UK. [88]Department of Neurology, Donders Center for Medical Neuroscience, Radboud University Medical Center, Nijmegen, The Netherlands. [89]Los Angeles County Department of Public Health, Los Angeles, CA, USA. [90]Institute of Biomedicine, Department of Laboratory Medicine, the Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden. [91]Department of Clinical Genetics and Genomics, Sahlgrenska University Hospital, Gothenburg, Sweden. [92]Department of Health Care Administration and Management, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan. [93]Department of Biology, Brody School of Medicine Center for Health Disparities, East Carolina University, Greenville, NC, USA. [94]Department of Neurology and Geriatric Research and Education Clinical Center, VA Maryland Health Care System, Baltimore, MD, USA. [95]Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Monash University, Melbourne, Victoria, Australia. [96]Intramural Research Program, National Institute on Aging, NIH, Baltimore, MD, USA. [97]Department of Neurology, Korea University Guro Hospital, Seoul, Republic of Korea. [98]Department of Computational Biology, University of Lausanne, Lausanne, Switzerland. [99]Swiss Institute of Bioinformatics, Lausanne, Switzerland. [100]University Center for Primary Care and Public Health, Lausanne, Switzerland. [101]Department of Epidemiology and Biostatistics, School of Public Health, Peking University Health Science Center, Beijing, China. [102]Stroke Research Group, Department of Clinical Neurosciences, University of Cambridge, Cambridge, UK. [103]Institute of Human Genetics, Technical University of Munich, Munich, Germany. [104]Institute of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. [105]Department of Medicine, University of Maryland School of Medicine, Baltimore, MD, USA. [106]Geriatrics Research and Education

Clinical Center, Baltimore Veterans Administration Medical Center, Baltimore, MD, USA. [107]Institute of Clinical Epidemiology and Biometry, University of Würzburg, Würzburg, Germany. [108]Department of Neurology, University Hospital Würzburg, Würzburg, Germany. [109]The MIND Center, University of Mississippi Medical Center, Jackson, MS, USA. [110]Department of Clinical Biochemistry, Copenhagen University Hospital—Herlev and Gentofte, Copenhagen, Denmark. [111]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark. [112]College of Medicine Nursing and Health Science, NUI Galway, Galway, Ireland. [113]Department of Genome Informatics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. [114]Laboratory for Systems Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. [115]Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Suita, Japan. [116]Integrated Frontier Research for Medical Science Division, Institute for Open and Transdisciplinary Research Initiatives, Osaka University, Suita, Japan. [117]Center for Infectious Disease Education and Research (CiDER), Osaka University, Suita, Japan. [118]Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, CA, USA. [119]Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. [120]Institute for Medical Information Processing, Biometry and Epidemiology, Ludwig Maximilian University Munich, Munich, Germany. [121]German Centre for Cardiovascular Research (DZHK), partner site Munich, Munich, Germany. [122]Department of Epidemiology, University of Washington, Seattle, WA, USA. [123]Department of Health Systems and Population Health, University of Washington, Seattle, WA, USA. [124]McCance Center for Brain Health, Massachusetts General Hospital, Boston, MA, USA. [125]Department of Neurology, University of Miami Miller School of Medicine, Miami, FL, USA. [126]Evelyn F. McKnight Brain Institute, Gainesville, FL, USA. [127]Division of Cardiology, Department of Medicine, Columbia University, New York, NY, USA. [128]Stroke Unit, Department of Neurology, Oslo University Hospital, Oslo, Norway. [129]Research and Development, The Norwegian Air Ambulance Foundation, Oslo, Norway. [130]Glenn Biggs Institute for Alzheimer's and Neurodegenerative Diseases, University of Texas Health Sciences Center, San Antonio, TX, USA. [131]Framingham Heart Study, Framingham, MA, USA. [132]University Medicine Greifswald, Institute for Community Medicine, SHIP/KEF, Greifswald, Germany. [133]Kaiser Permanente Washington Health Research Institute, Kaiser Permanente Washington, Seattle, WA, USA. [134]Department of Veterans Affairs Office of Research and Development, Seattle Epidemiologic Research and Information Center, Seattle, WA, USA. [135]Department of Neurology, University of Pennsylvania, Philadelphia, PA, USA. [136]Atlanta VA Health Care System, Decatur, GA, USA. [137]Department of Epidemiology, Emory University Rollins School of Public Health, Atlanta, GA, USA. [138]Department of Clinical Neuroscience, Institute of Neuroscience and Physiology, Sahlgrenska Unviersity Hospital, Gothenburg, Sweden. [139]Department of Internal Medicine, Section of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, The Netherlands. [140]Department of Cardiology, Leiden University Medical Center, Leiden, The Netherlands. [141]Division Heart & Lungs, Department of Cardiology, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands. [142]Department of Neurology and Neurosurgery, University of Tartu, Tartu, Estonia. [143]Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA, USA. [144]Department of Public Health and Clinical Medicine, Umeå University, Umeå, Sweden. [145]Department of Medicine, Division of Cardiovascular Disease, Emory University School of Medicine, Atlanta, GA, USA. [146]Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA. [147]Division of Genome Science, Department of Precision Medicine, National Institute of Health, Cheongju, Republic of Korea. [148]Research Unit Molecular Epidemiology, Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. [149]Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Germany. [150]German Center for Neurodegenerative Diseases (DZNE), site Rostock/Greifswald, Rostock, Germany. [151]Netherlands Heart Institute, Utrecht, The Netherlands. [152]Einthoven Laboratory for Experimental Vascular Medicine, LUMC, Leiden, The Netherlands. [153]Department of Clinical Genetics, Department of Radiology and Nuclear Medicine, Erasmus MC, Rotterdam, The Netherlands. [154]Latin American Brain Health (BrainLat), Universidad Adolfo Ibáñez, Santiago, Chile. [155]Department of Epidemiology and Population Health, Albert Einstein College of Medicine, New York, NY, USA. [156]Department of Clinical Biochemistry and Pharmacology, Odense University Hospital, Odense, Denmark. [157]Department of Neurology, University of Virginia, Charlottesville, VA, USA. [158]Department of Public Health Science, University of Virginia, Charlottesville, VA, USA. [159]McGill Genome Centre, Montreal, Quebec, Canada. [160]Longitudinal Studies Section, Translational Gerontology Branch, National Institute on Aging, Baltimore, MD, USA. [161]Geisinger Neuroscience Institute, Geisinger Health System, Danville, PA, USA. [162]Department of Neurology, College of Medicine, The Pennsylvania State University, State College, PA, USA. [163]Klinik für Neurologie, Carl von Ossietzky University of Oldenburg, Oldenburg, Germany. [164]Comprehensive Heart Failure Center, University Hospital Würzburg, Würzburg, Germany. [165]Clinical Trial Center, University Hospital Würzburg, Würzburg, Germany. [166]Department of Neurology, University of

Washington, Seattle, WA, USA. [167]Institute for Health Promotion, Graduate School of Public Health, Yonsei University, Seoul, Republic of Korea. [168]Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, Ontario, Canada. [169] Population Health Research Institute, David Braley Cardiac, Vascular and Stroke Research Institute, Hamilton, Ontario, Canada. [170]Department of Surgery and Department of Genetics, University of Pennsylvania, Philadelphia, PA, USA. [171]Corporal Michael Crescenz VA Medical Center, Philadelphia, PA, USA. [172]The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA, USA. [173]Department of Neurology, Brigham and Women's Hospital, Boston, MA, USA. [174]Division of Medicine, Turku University Hospital, Turku, Finland. [175]Brown Foundation Institute of Molecular Medicine, McGovern Medical School, University of Texas Health Science Center at Houston, Houston, TX, USA. [176]Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, USA. [177]Department of Medical Imaging, Chung Shan Medical University Hospital, Taichung, Taiwan. [178]Department of Neurology, Boston University School of Medicine, Boston, MA, USA. [179]Department of Medicine, University of Ibadan, Ibadan, Nigeria. [180]Munich Cluster for Systems Neurology, Munich, Germany. [181]German Center for Neurodegenerative Diseases (DZNE), Munich, Germany. [182]Department of Neurology, Institute for Neurodegenerative Diseases, CHU de Bordeaux, Bordeaux, France.

EADB Consortium

Céline Bellenguez[1], Benjamin Grenier-Boley[1], Sven van der lee[2,3], Iris Jansen[2,4], Rafael Campos-Martin[5], Peter A. Holmans[6], Luca Kleineidam[5,7,8], Oliver Peters[9], Anja Schneider[10], Martin Dichgans[11,12,13], Dan Rujescu[14], Norbert Scherbaum[15], Jürgen Deckert[16], Steffi Riedel-Heller[17], Lucrezia Hausner[18], Laura Molina Porcel[19], Emrah Düzel[20], Timo Grimmer[21], Jens Wiltfang[22,23,24], Stefanie Heilmann-Heimbach[25], Michael Wagner[26,27], Susanne Moebus[28], Thomas Tegos[29], Nikolaos Scarmeas[30,31], Oriol Dols[32], Fermin Moreno[32], Jordi Pérez-Tur[32], María J. Bullido[32], Pau Pastor[33], Raquel Sánchez-Valle[34], Victoria Álvarez[35], Hilkka Soininen[36], Sami Heikkinen[37], Alexandre de Mendonça[38], Shima Mehrabian[39], Jakub Hort[40,41], Martin Vyhnalek[40,41], Nicolai Sandau[42], Jesper Qvist-Thomassen[42], Caroline Graff[43], Goran Papenberg[44], Julie Williams[45], Philippe Amouyel[46], Anne Boland[47], Jean-François Deleuze[47], Gael Nicolas[48], Florence Pasquier[49], Olivier Hanon[50], Edna Grünblatt[51,52,53], Julius Popp[54,55,56], Giacomina Rossi[57], Daniela Galimberti[58,59], Beatrice Arosio[60,61], Patrizia Mecocci[62,63], Vincenzo Solfrizzi[64], Lucilla Parnetti[65], Alessio Squassina[66], Lucio Tremolizzo[67], Barbara Borroni[68], Benedetta Nacmias[69,70], Marco Spallazzi[71], Davide Seripa[72], Innocenzo Rainero[73], Antonio Daniele[74,75], Fabrizio Piras[76], Carlo Masullo[46], Emilio Alarcón-Martín[77,78], Daniel Alcolea[79,32], Ignacio Alvarez[80,81], Tsolaki Anthoula[29,82], Ildebrando Appollonio[83,84], Marina Arcaro[85], Silvana Archetti[86], Alfonso Arias Pastor[87,88], Henri Bailly[50], Nerisa Banaj[89], Miquel Baquero[90], Ana Belén Pastor[91], Luisa Benussi[92], Claudine Berr[93], Céline Besse[47], Valentina Bessi[94,95], Giuliano Binetti[92,96], Alessandra Bizarro[97], Rafael Blesa[79,32], Silvia Boschi[73], Paola Bossù[98], Catherine Bresner[6], Keeley J. Brookes[99], Dolores Buiza-Rueda[32,100], Katharina Bûrger[101,102], Vanessa Burholt[103,104], Miguel Calero[32,91,105], Geneviève Chene[106,107], Ángel Carracedo[108,109], Roberta Cecchetti[110], Laura Cervera-Carles[79,32], Camille Charbonnier[111], Caterina Chillotti[112], Simona Ciccone[113], Jurgen A.H.R. Claassen[114], Jordi Clarimon[79,32], Christopher Clark[115], Elisa Conti[83], Anaïs Corma-Gómez[116], Emanuele COSTANTINI[117], Carlo Custodero[118], Delphine Daian[47], Efthimios Dardiotis[119], Jean-François Dartigues[120], Peter P de Deyn[121], Teodoro del Ser[122], Nicola Denning[123], Janine Diehl-Schmid[124], Mónica Diez-Fairen[80,81], Paolo Dionigi Rossi[113], Emmanuelle Duron[50], Sebastiaan Engelborghs[125,126,127,128], Valentina Escott-Price[123,6], Michael Ewers[101,102], Tagliavini Fabrizio[129], Sune Fallgaard Nielsen[130], Lucia Farotti[131], Chiara Fenoglio[132], Marta Fernández-Fuertes[116], Catarina B Ferreira[38], Evelyn Ferri[113], Bertrand Fin[47], Peter Fischer[133], Klaus Fließbach[134,8], Juan Fortea[79,32], Tatiana M. Foroud[135], Silvia Fostinelli[92], Nick C. Fox[136], Emlio Franco-Macías[137], Ana Frank-García[32,138,139], Lutz Froelich[140], Jose-Maria García-Alberca[32,141], Sebastian Garcia-Madrona[142], Guillermo Garcia-Ribas[142], Ina Giegling[143], Giaccone Giorgio[129], Oliver Goldhardt[124], Antonio González-Pérez[144], Giulia Grande[44], Emma Green[145], Edna Grünblatt[51,52,53], Tamar Guetta-Baranes[146], Annakaisa Haapasalo[147], Georgios Hadjigeorgiou[148], Harald Hampel[149,150], John Hardy[151], Annette M. Hartmann[143], Janet Harwood[6], Seppo Helisalmi[152,153], Michael T. Heneka[7,154], Martin J. Herrmann[155], Per Hoffmann[25], Clive Holmes[156], Raquel Huerto Vilas[87,88], Marc Hulsman[2,3], Geert Jan Biessels[157], Charlotte Johansson[158,159], Lena Kilander[160], Anne Kinhult Ståhlbom[158,159], Miia KIVIPELTO[161,162,163,164], Anne Koivisto[152,165,166], Johannes Kornhuber[167], Mary H. Kosmidis[168], Carmen Lage[32,169], Erika J Laukka[44,170], Alessandra Lauria[97], Jenni LEHTISALO[152,171], Ondrej LERCH[40,41], Alberto Lleó[79,32], Adolfo Lopez Lopez De Munain[32,172], Seth Love[173], Malin Löwemark[160], Lauren Luckcuck[6], Juan Macías[116], Catherine A. MacLeod[174], Wolfgang Maier[175,8], Francesca MANGIALASCHE[161], Spallazzi Marco[176], Rachel Marshall[6], Angel Martín[32,138,139], Carmen Martínez Rodríguez[177], Simon Mead[178], Miguel Medina[32,91], Alun Meggy[123], Silvia

Mendoza[141], Manuel Menéndez-González[177], Pablo Mir[32,100], Merel Mol[179], Kevin Morgan[180], Markus M Möthen[25], Tiia NGANDU[171], Børge G. Nordestgaard[181,182], Robert Olaso[47], Michela Orsini[117], Alessandro Padovani[183], Caffarra Paolo[184], Pierre Pericard[185], Yolande A.L. Pijnenburg[2], Juan A Pineda[116], Gerard Piñol-Ripoll[87,88], Claudia Pisanu[186], Thomas Polak[155], Danielle Posthuma[4], Josef Priller[187,188], Olivier Quenez[111], Inés Quintela[108], Alberto Rábano[91,189,32], Luis M Real[116,190], Marcel J.T. Reinders[191], Peter Riederer[192], Eloy Rodriguez-Rodriguez[169,32], Irene Rosas Allende[193,35], Jose Luis Royo[194], Elisa Rubino[195], María E Sáez[144], Paraskevi Sakka[196], María B Sánchez-Arjona[137], Florentino Sanchez-Garcia[197], Raquel Sánchez-Valle[198], Michela Scamosci[110], Elio Scarpini[85,132], Martin Scherer[199], Matthias Schmid[154,200], Jonathan M. Schott[136], Olivia Skrobot[173], Alina Solomon[152,161], Sandro Sorbi[94,70], Annika Spottke[154,201], Niccolo Tesí[2,3], Tegos Thomas[29], Latchezar Traykov[39], Anne Tybjærg-Hansen[202], Andre Uitterlinden[203], Abbe Ullgren[158], Christine van Broeckhoven[204,205,206], Jasper van Dongen[207,204,205], Rik Vandenberghe[208,209], Jean-Sébastien Vidal[50], Jonathan Vogelgsang[22,210], Michael Wagner[134,154], David Wallon[211], Leonie Weinhold[200], Gill Windle[174], Bob Woods[174], Mary Yannakoulia[212], Miren Zulaica[32,213], Carolina Dalmasso[214], Yolande AL Pijnenburg[2], Henne Holstege[2], John van Swieten[179], Inez Ramakers[215], Frans Verhey[215], Aad van der Lugt[216], Philip Scheltens[2], Frank Jessen[217,27,218], Patrick Kehoe[219], Tsolaki Magda[29,220], Ruth Frikke-Schmidt[42,182], Pascual Sánchez-Juan[32], Roberta Ghidoni[221], Kristel Sleegers[207,222], Vilmantas Giedraitis[223], Martin Ingelsson[224,225,226], Mikko Hiltunen[37], Rebecca Sims[45], Alfredo Ramirez[214,26,227,228,27], Jean-Charles Lambert[1]

[1]Univ. Lille, Inserm, CHU Lille, Institut Pasteur de Lille, U1167-RID-AGE facteurs de risque et déterminants moléculaires des maladies liés au vieillissement, Lille, France, [2]Alzheimer Center Amsterdam, Department of Neurology, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam UMC, Amsterdam, The Netherlands, [3]Section Genomics of Neurdegenerative Diseases and Aging, Department of Human Genetics Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam UMC, Amsterdam, The Netherlands, [4]Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, Vrije University, Amsterdam, The Netherlands., [5]Division of Neurogenetics and Molecular Psychiatry, Department of Psychiatry and Psychotherapy, University of Cologne, Medical Faculty, 50937 Cologne, Germany., [6]MRC Centre for Neuropsychiatric Genetics and Genomics, , Division of Psychological Medicine and Clinical Neuroscience, School of Medicine, Cardiff University, Cardiff, UK, [7]Department of Neurodegeneration and Geriatric Psychiatry, University of Bonn, Bonn, Germany., [8]German Center for Neurodegenerative Diseases (DZNE Bonn), Bonn, Germany, [9]Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Institute of Psychiatry and Psychotherapy, Hindenburgdamm 30, 12203 Berlin, Germany, [10]Department of Cognitive Disorders and Old Age Psychiatry, University Hospital Bonn, Venusberg-Campus 1, 53127 Bonn, Germany, [11]Institute for Stroke and Dementia Research (ISD), University Hospital, LMU Munich, Munich, Germany., [12]German Center for Neurodegenerative Diseases (DZNE), Munich, Germany, [13]Munich Cluster for Systems Neurology (SyNergy), Munich, Germany, [14]Martin-Luther-University Halle-Wittenberg, University Clinic and Outpatient Clinic for Psychiatry, Psychotherapy and Psychosomatics, Halle (Saale), Germany, [15]Department of Psychiatry and Psychotherapy, LVR-Klinikum Essen, University of Duisburg-Essen, Germany Medical Faculty, Germany, [16]Department of Psychiatry, Psychosomatics and Psychotherapy, Center of Mental Health, University Hospital of Würzburg, Germany, [17]Institute of Social Medicine, Occupational Health and Public Health, University of Leipzig, 04103 Leipzig, Germany., [18]Department of Geriatric Psychiatry, Central Institute for Mental Health Mannheim, Faculty Mannheim, University of Heidelberg, Germany, [19]Alzheimer's disease and other cognitive disorders Unit. Neurology Department, Hospital Clinic , Barcelona, Spain, [20]Institute of Cognitive Neurology and Dementia Research (IKND), Otto-von-Guericke University, Magdeburg, Germany, [21]Center for Cognitive Disorders, Department of Psychiatry and Psychotherapy, Technical University of Munich, School of Medicine, Munich, Germany, [22]Department of Psychiatry and Psychotherapy, University Medical Center Goettingen, Goettingen, Germany, [23]German Center for Neurodegenerative Diseases (DZNE), Goettingen, Germany, [24]Medical Science Department, iBiMED, Aveiro, Portugal, [25]Institute of Human Genetics, University of Bonn, School of Medicine & University Hospital Bonn, Bonn, Germany, [26]Department of Cognitive Disorders and Old Age Psychiatry, University Hospital Bonn, Medical Faculty, Bonn, Germany, [27]German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany, [28]Institute for Urban Public Health, University Hospital of University Duisburg-Essen, Essen, Germany., [29]1st Department of Neurology, Medical school, Aristotle University of Thessaloniki, Thessaloniki, Makedonia, Greece, [30]Taub Institute for Research in Alzheimer's Disease and the Aging Brain, The Gertrude H. Sergievsky Center, Depatment of Neurology, Columbia University, New York, NY, [31]1st Department of Neurology, Aiginition Hospital, National and Kapodistrian University of Athens, Medical School, Greece, [32]CIBERNED, Network Center for Biomedical Research in

Neurodegenerative Diseases, National Institute of Health Carlos III, Madrid, Spain, [33]Memory Disorders Unit, Department of Neurology, Hospital Universitari Mutua de Terrassa, Terrassa, Barcelona, Spain, [34]Alzheimer's disease and other cognitive disorders unit. Service of Neurology. Hospital Clínic of Barcelona. Institut d'Investigacions Biomèdiques August Pi i Sunyer, University of Barcelona, Barcelona, Spain, [35]Instituto de Investigación Sanitaria del Principado de Asturias (ISPA), [36]Institute of Clinical Medicine - Neurology, University of Eastern Finland, Finland, [37]Institute of Biomedicine, University of Eastern Finland, Finland, [38]Faculty of Medicine, University of Lisbon, Portugal, [39]Clinic of Neurology, UH "Alexandrovska", Medical University - Sofia, Sofia, Bulgaria, [40]Memory Clinic, Department of Neurology, Charles University, 2nd Faculty of Medicine and Motol University Hospital, Czech Republic, [41]International Clinical Research Center, St. Anne's University Hospital Brno, Brno, Czech Republic, [42]Department of Clinical Biochemistry, Copenhagen University Hospital - Rigshospitalet, Copenhagen, Denmark, [43]Unit for Hereditary Dementias, Theme Aging, Karolinska University Hospital-Solna, 171 64 Stockholm Sweden, [44]Aging Research Center, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet and Stockholm University, Stockholm, Sweden, [45]Division of Psychological Medicine and Clinical Neuroscience, School of Medicine, Cardiff University, Wales, UK, [46]Institute of Neurology, Catholic University of the Sacred Heart, Rome, Itlay, [47]Université Paris-Saclay, CEA, Centre National de Recherche en Génomique Humaine, 91057, Evry, France, [48]Normandie Univ, UNIROUEN, Inserm U1245 and Rouen University Hospital, Department of Genetics and CNR-MAJ, F 76000, Normandy Center for Genomic and Personalized Medicine, Rouen, France, [49]Univ Lille Inserm 1171, CHU Clinical and Research Memory Research Centre (CMRR) of Distalz Lille France., [50]Université de Paris, EA 4468, APHP, Hôpital Broca, Paris, France, [51]Department of Child and Adolescent Psychiatry and Psychotherapy, University Hospital of Psychiatry Zurich, University of Zurich, Zurich, Switzerland, [52]Neuroscience Center Zurich, University of Zurich and ETH Zurich, Switzerland, [53]Zurich Center for Integrative Human Physiology, University of Zurich, Switzerland, [54]Old Age Psychiatry, Department of Psychiatry, Lausanne University Hospital, Lausanne, Switzerland, [55]Department of Geriatric Psychiatry, University Hospital of Psychiatry Zürich, Zürich, Switzerland, [56]Institute for Regenerative Medicine, University of Zürich, Switzerland, [57]Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy, [58]Neurodegenerative Diseases Unit, Fondazione IRCCS Ca' Granda, Ospedale Policlinico, Milan, IT, [59]Dept. of Biomedical, Surgical and Dental Sciences, University of Milan, Milan, IT, [60]Department of Clinical Sciences and Community Health, University of Milan, 20122 Milan, Italy, [61]Geriatric Unit, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, 20122 Milan, Italy, [62]Institute of Gerontology and Geriatrics,Department of Medicine and Surgery, University of Perugia, Italy, [63]Division of Clinical Geriatrics, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden, [64]Interdisciplinary Department of Medicine, Geriatric Medicine and Memory Unit, University of Bari "A. Moro, Bari, Italy, [65]Centre for Memory Disturbances, Lab of Clinical Neurochemistry, Section of Neurology, University of Perugia, Perugia, Italy, [66]Department of Biomedical Sciences, Section of Neuroscience and Clinical Pharmacology,University of Cagliari, Cagliari, Italy., [67]Neurology Unit, "San Gerardo" hospital, Monza and University of Milano-Bicocca, Milan, Italy, [68]Cognitive and Behavioural Neurology, Department of Continuity of Care and Frailty, ASST Spedali Civili Brescia, Italy, [69]Department of Neuroscience, Psychology, Drug Research and Child Health University of Florence, Florence, Italy, [70]IRCCS Fondazione Don Carlo Gnocchi, Florence, Italy, [71]Department of Medicine and Surgery, Unit of Neurology, University-Hospital of Parma, Parma, Italy, [72]Department of Hematology and Stem Cell Transplant, Vito Fazzi Hospital, Lecce, Italy, [73]Department of Neuroscience "Rita Levi Montalcini", University of Torino, Torino, Italy, [74]Department of Neuroscience, Università Cattolica del Sacro Cuore, Rome, Italy., [75]Neurology Unit, IRCCS Fondazione Policlinico Universitario A. Gemelli, Rome, Italy., [76]Laboratory of Neuropsychiatry, IRCCS Santa Lucia Foundation, Rome, Italy, [77]Research Center and Memory clinic Fundació ACE, Institut Català de Neurociències Aplicades, Universitat Internacional de Catalunya, Barcelona, Spain, [78]Department of Surgery, Biochemistry and Molecular Biology, School of Medicine, University of Málaga, Málaga, Spain., [79]Department of Neurology, II B Sant Pau, Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, Barcelona, Spain., [80]Fundació Docència i Recerca MútuaTerrassa and Movement Disorders Unit, Department of Neurology, University Hospital MútuaTerrassa, Terrassa 08221, Barcelona, Spain, [81]Memory Disorders Unit, Department of Neurology, Hospital Universitari Mutua de Terrassa, Terrassa, Barcelona, Spain., [82]Alzheimer Hellas, Thessaloniki, Makedonia, Greece, [83]School of Medicine and Surgery, University of Milano-Bicocca, Italy, [84]Neurology Unit, "San Gerardo" hospital, Monza, Italy, [85]Fondazione IRCCS Ca' Granda, Ospedale Policlinico, Milan, Italy, [86]Department of Laboratory Diagnostics, III Laboratory of Analysis, Brescia Hospital, Brescia, Italy, [87]Unitat Trastorns Cognitius, Hospital Universitari Santa Maria de Lleida, Lleida, Spain, [88]Institut de Recerca Biomedica de Lleida (IRBLLeida), Lleida, Spain, [89]Laboratory of Neuropsychiatry, Department of Clinical and Behavioral Neurology, IRCCS Santa Lucia Foundation, Rome, Italy, [90]Servei de Neurologia, Hospital Universitari i Politècnic La Fe, Valencia, Spain., [91]CIEN Foundation/Queen Sofia Foundation

Alzheimer Center, [92]Molecular Markers Laboratory, IRCCS Istituto Centro San Giovanni di Dio Fatebenefratelli, Brescia, [93]Univ. Montpellier, Inserm U1061, Neuropsychiatry: epidemiological and clinical research, PSNREC, Montpellier, France, [94]Department of Neuroscience, Psychology, Drug Research and Child Health University of Florence, Florence Italy, [95]Azienda Ospedaliero-Universitaria Careggi, , Florence, Italy, [96]MAC - Memory Clinic, IRCCS Istituto Centro San Giovanni di Dio Fatebenefratelli, Brescia, [97]Geriatrics Unit Fondazione Policlinico A. Gemelli IRCCS, Rome , Italy, [98]Experimental Neuro-psychobiology Laboratory,Department of Clinical and Behavioral Neurology, IRCCS Santa Lucia Foundation, Rome, Italy, [99]Biosciences, School of Science and Technology, Nottingham Trent University, Nottingham UK, [100]Unidad de Trastornos del Movimiento, Servicio de Neurología y Neurofisiología. Instituto de Biomedicina de Sevilla (IBiS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Seville, Spain, [101]Institute for Stroke and Dementia Research, Klinikum der Universität München, Ludwig-Maximilians-Universität LMU, Munich, Germany., [102]German Center for Neurodegenerative Diseases (DZNE, Munich), Munich, Germany., [103]Faculty of Medical & Health Sciences, University of Auckland, New Zealand, [104]Wales Centre for Ageing & Dementia Research, Swansea University, Wales, New Zealand, [105]UFIEC, Instituto de Salud Carlos III, , Madrid, Spain, [106]Inserm, Bordeaux Population Health Research Center, UMR 1219, Univ. Bordeaux, ISPED, CIC 1401-EC, Univ Bordeaux,Bordeaux, France, [107]CHU de Bordeaux, Pole santé publique, Bordeaux, France, [108]Grupo de Medicina Xenómica, Centro Nacional de Genotipado (CEGEN-PRB3-ISCIII). Universidade de Santiago de Compostela, Santiago de Compostela, Spain., [109]Fundación Pública Galega de Medicina Xenómica- CIBERER-IDIS, University of Santiago de Compostela, Santiago de Compostela, Spain., [110]Institute of Gerontology and Geriatrics, Department of Medicine and Surgery, University of Perugia Perugia (Italy), [111]Normandie Univ, UNIROUEN, Inserm U1245 and CHU Rouen, Department of Genetics and CNR-MAJ, Rouen, France, [112]Unit of Clinical Pharmacology, University Hospital of Cagliari, Cagliari, Italy, [113]Geriatic Unit, Fondazione Cà Granda, IRCCS Ospedale Maggiore Policlinico, Milan, Italy, [114]Radboudumc Alzheimer Center, Department of Geriatrics, Radboud University Medical Center, Nijmegen, the Netherlands, [115]Institute for Regenerative Medicine, University of Zürich, Schlieren, Switzerland, [116]Unidad Clínica de Enfermedades Infecciosas y Microbiología. Hospital Universitario de Valme, Sevilla, Spain, [117]Department of Neuroscience, Catholic University of Sacred Heart, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy, [118]University of Bari, "A. Moro", [119]School of Medicine, University of Thessaly, Larissa, Greece, [120]University Bordeaux, Inserm, Bordeaux Population Health Research Center, France, [121]Department of Neurology, University Medical Center Groningen, the Netherlands, [122]Department of Neurology/CIEN Foundation/Queen Sofia Foundation Alzheimer Center, [123]UKDRI@ Cardiff, School of Medicine, Cardiff University, Cardiff, UK, [124]Technical University of Munich, School of Medicine, Klinikum rechts der Isar, Department of Psychiatry and Psychotherapy, [125]Center for Neurosciences, Vrije Universiteit Brussel (VUB), Brussels, Belgium, [126]Reference Center for Biological Markers of Dementia (BIODEM), Institute Born-Bunge, University of Antwerp, Antwerp, Belgium, [127]Institute Born-Bunge, University of Antwerp, Antwerp, Belgium, [128]Department of Neurology, UZ Brussel, Brussels, Belgium, [129]Fondazione IRCCS, Istituto Neurologico Carlo Besta, Milan Italy, [130]Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Herlev Denmark, [131]Centre for Memory Disturbances, Lab of Clinical Neurochemistry, Section of Neurology, University of Perugia, Italy, [132]University of Milan, Milan, Italy, [133]Department of Psychiatry, Social Medicine Center East- Donauspital, Vienna, Austria, [134]Department of Neurodegeneration and Geriatric Psychiatry, University of Bonn, 53127 Bonn, Germany., [135]Department of Medical and Molecular Genetics, Indiana University, Indianapolis, Indiana, USA, [136]Dementia Research Centre, UCL Queen Square Institute of Neurology, London, United Kingdom, [137]Unidad de Demencias, Servicio de Neurología y Neurofisiología. Instituto de Biomedicina de Sevilla (IBiS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Seville, Spain, [138]Instituto de Investigacion Sanitaria 'Hospital la Paz' (IdIPaz), Madrid, Spain, [139]Hospital Universitario la Paz, Madrid, Spain, [140]Department of geriatric Psychiatry, Central Institute for Mental Health, Mannheim, University of Heidelberg, Germany, [141]Alzheimer Research Center & Memory Clinic, Andalusian Institute for Neuroscience, Málaga, Spain., [142]Hospital Universitario Ramon y Cajal, IRYCIS, Madrid, [143]Department of Psychiatry and Psychotherapy, Medical University of Vienna, Vienna, Austria, [144]CAEBI, Centro Andaluz de Estudios Bioinformáticos, Sevilla, Spain., [145]Institute of Public Health, University of Cambridge, UK, [146]Human Genetics, School of Life Sciences, Life Sciences Building, University Park, University of Nottingham, Nottingham, UK, [147]A.I Virtanen Institute for Molecular Sciences, University of Eastern Finland, Kuopio, Finland, [148]Department of Neurology, Medical School, University of Cyprus, Cyprus, [149]Sorbonne University, GRC n° 21, Alzheimer Precision Medicine Initiative (APMI), AP-HP, Pitié-Salpêtrière Hospital, Boulevard de l'hôpital, F-75013, Paris, France, [150]Eisai Inc., Neurology Business Group, 100 Tice Blvd, Woodcliff Lake, NJ 07677, USA, [151]Reta Lila Weston Research Laboratories, Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK., [152]Insitute of Clinical Medicine - Neurology, University of Eastern, Kuopio, Finland, [153]Institute of Clinical

Medicine – Internal Medicine, University of Eastern Finland, Kuopio, Finland, [154]German Center for Neurodegenerative Diseases (DZNE, Bonn), Bonn, Germany., [155]Department of Psychiatry, Psychosomatics and Psychotherapy, Center of Mental Health, University Hospital, Wuerzburg, [156]Clinical and Experimental Science, Faculty of Medicine, University of Southampton, Southampton, UK., [157]Department of Neurology, UMC Utrecht Brain Center, Utrecht, the Netherlands, [158]Karolinska Institutet, Center for Alzheimer Research, Department NVS, Division of Neurogeriatrics, Stockholm, Sweden, [159]Unit for Hereditary dementias, Karolinska University Hospital-Solna, Stockholm, Sweden, [160]Dept.of Public Health and Carins Sciences / Geriatrics, Uppsala University, [161]Division of Clinical Geriatrics, Center for Alzheimer Research, Care Sciences and Society (NVS), Karolinska Institutet, Stockholm, Sweden, [162]Institute of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, Finland, [163]Neuroepidemiology and Ageing Research Unit, School of Public Health, Imperial College London, London, United Kingdom, [164]Stockholms Sjukhem, Research & Development Unit, Stockholm, Sweden, [165]Department of Neurology, Kuopio University Hospital, Kuopio, Finland, [166]Department of Neurosciences, University of Helsinki and Department of Geriatrics, Helsinki University Hospital, Helsinki, Finland, [167]Department of Psychiatry and Psychotherapy, Universitätsklinikum Erlangen, and Friedrich-Alexander Universität Erlangen-Nürnberg, Erlangen, Germany., [168]Laboratory of Cognitive Neuroscience, School of Psychology, Aristotle University of Thessaloniki, Thessaloniki, Greece, [169]Neurology Service, Marqués de Valdecilla University Hospital (University of Cantabria and IDIVAL), Santander, Spain., [170]Stockholm Gerontology Research Center, Stockholm, Sweden, [171]Public Health Promotion Unit, Finnish Institute for Health and Welfare, Helsinki, Finland, [172]Department of Neurology. Hospital Universitario Donostia. OSAKIDETZA-Servicio Vasco de Salud, San Sebastian, Spain, [173]Translational Health Sciences, Bristol Medical School, University of Bristol, Bristol, BS16 1LE, UK, [174]School of Health Sciences, Bangor University, UK, [175]Department of Neurodegenerative Diseases and Geriatric Psychiatry, University Hospital Bonn, Bonn, Germany, [176]Unit of Neurology, University of Parma and AOU, Parma, Italy, [177]Servicio de Neurología HOspital Universitario Central de Asturias- Oviedo and Instituto de Investigación Biosanitaria del Principado de Asturias, Oviedo, Spain, [178]MRC Prion Unit at UCL, UCL Institute of Prion Diseases, London, W1W 7FF, [179]Department of Neurology, ErasmusMC, [180]Human Genetics, School of Life Sciences, University of Nottingham, UK NG7 2UH, [181]Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Herlev, Denmark, [182]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark, [183]Centre for Neurodegenerative Disorders, Department of Clinical and Experimental Sciences, University of Brescia, Brescia, Italy, [184]DIMEC, University of Parma, Parma, Italy, [185]Univ. Lille, CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, US 41-UMS 2014-PLBS, bilille, Lille, France., [186]Department of Biomedical Sciences, University of Cagliari, Italy, [187]German Center for Neurodegenerative Diseases (DZNE), Berlin, Germany., [188]Department of Neuropsychiatry and Laboratory of Molecular Psychiatry, Charité, Charitéplatz 1, 10117 Berlin, Germany, [189]BT-CIEN, [190]Depatamento de Especialidades Quirúrgicas, Bioquímica e Inmunología. Facultad de Medicina. Universidad de Málaga. Málaga, Spain, [191]Delft Bioinformatics Lab, Delft University of Technology, Delft, The Netherlands, [192]Center of Mental Health, Clinic and Policlinic of Psychiatry, Psychosomatics and Psychotherapy, University Hospital of Würzburg, Wuerzburg, Germany, [193]Laboratorio de Genética. Hospital Universitario Central de Asturias, Oviedo, Spain, [194]Departamento de Especialidades Quirúrgicas, Bioquímicas e Inmunología, School of Medicine, University of Málaga, Málaga, Spain., [195]Department of Neuroscience and Mental Health, AOU Città della Salute e della Scienza di Torino, Torino, Italy, [196]Athens Association of Alzheimer's disease and Related Disorders, Athens, Greece, [197]Department of Immunology, Hospital Universitario Doctor Negrín, Las Palmas de Gran Canaria, Spain., [198]Neurology department-Hospital Clínic, IDIBAPS, Universitat de Barcelona, Barcelona, Spain., [199]Department of Primary Medical Care, University Medical Centre Hamburg-Eppendorf, 20246 Hamburg, Germany., [200]Institute of Medical Biometry, Informatics and Epidemiology, University Hospital of Bonn, Bonn, Germany., [201]Department of Neurology, University of Bonn, Bonn, Germany., [202]Department of Clinical Biochemistry, Rigshospitalet, Copenhagen, Denmark, [203]Department of Internal medicine and Biostatistics, ErasmusMC, [204]Laboratory of Neurogenetics, Institute Born - Bunge, Antwerp, Belgium, [205]Department of Biomedical Sciences, University of Antwerp, Neurodegenerative Brain Diseases Group, Center for Molecular Neurology, VIB, Antwerp, Belgium, [206]Neurodegenerative Brain Diseases Group, VIB Center for Molecular Neurology, VIB, Antwerp, Belgium, [207]Complex Genetics of Alzheimer's Disease Group, VIB Center for Molecular Neurology, VIB, Antwerp, Belgium, [208]Laboratory for Cognitive Neurology, Department of Neurosciences, University of Leuven, Belgium, [209]Neurology Department, University Hospitals Leuven, Leuven, Belgium, [210]Department of Psychiatry, Harvard Medical School, McLean Hospital, Belmont, MA, USA, [211]249. Normandie Univ, UNIROUEN, Inserm U1245, CHU Rouen, Department of Neurology and CNR-MAJ, F 76000, Normandy Center for Genomic and Personalized Medicine, Rouen, France, [212]Department of Nutrition and Diatetics, Harokopio University, Athens, Greece, [213]Neurosciences Area. Instituto Biodonostia. San Sebastian, Spain, [214]Division of Neurogenetics and Molecular Psychiatry, Department

of Psychiatry and Psychotherapy, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany, [215]Maastricht University, Department of Psychiatry & Neuropsychologie, Alzheimer Center Limburg, Maastricht, the Netherlands, [216]Department of Radiology&Nuclear medicine, ErasmusMC, [217]Department of Psychiatry and Psychotherapy, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany, [218]Cluster of Excellence Cellular Stress Responses in Aging-associated Diseases (CECAD), University of Cologne, Cologne, Germany, [219]Translational Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK, [220]Laboratory of Genetics, Immunology and Human Pathology, Faculty of Science of Tunis, University of Tunis El Manar, 2092 Tunis, Tunisia., [221]Molecular Markers Laboratory, IRCCS Istituto Centro San Giovanni di Dio Fatebenefratelli, Brescia, Italy, [222]Department of Biomedical Sciences, University of Antwerp, Antwerp, Belgium, [223]Dept.of Public Health and Carins Sciences / Geriatrics, Uppsala University, Sweden, [224]Dept. of Public Health and Caring Sciences / Geriatrics, Uppsala University, Uppsala, Sweden, [225]Krembil Brain Institute, University Health Network, Toronto, Ontario, Canada, [226]Tanz Centre for Research in Neurodegenerative Diseases, Departments of Medicine and Laboratory Medicine & Pathobiology, University of Toronto, Toronto, Ontario, Canada, [227]Department of Psychiatry & Glenn Biggs Institute for Alzheimer's and Neurodegenerative Diseases, San Antonio, TX, USA, [228]Cologne Excellence Cluster on Cellular Stress Responses in Aging-Associated Disease (CECAD), University of Cologne, Cologne, Germany.

Supplementary Tables

eTable 1. ICD-9 and ICD-10 codes to identify dementia cases in All of Us. 27

eTable 2. Input traits into iPRS-DEM model and associated GWAS references. 29

eTable 3. Best PRS for each individual trait in 3C-Bordeaux-Montpellier training set. Best model was selected on basis of lowest integrated brier score. Fine-Gray models are adjusted for age at baseline, age-squared , study centre, sex, the first 10 PCs, and dosage of APOE-ε4 and APOE-ε2. 30

eTable 4. Single-trait PRS Training weights derived from elastic net. Elastic-net was adjusted for age at baseline, age-squared , study centre, sex, the first 10 PCs, and dosage of APOE-ε4 and APOE-ε2. The PRS weight is derived by dividing the elastic-net coefficient by the empirical standard deviation. Optimal elastic-net parameters: α=0.0787; λ=0.001163819..... 31

eTable 5. Top 250 SNPs ordered by the absolute value of their iPRS-weight calculated across all single-trait PRS included in the integrative model. 32

eTable 6. Association of APOE, iPRS and AD-PRS with dementia or death in 3C-Dijon using Fine-Gray and Cox proportional hazard models. All models are adjusted for age at baseline (with a quadratic term), sex, 10 PCs. Each model with PRS is also adjusted for APOE-ε4 and APOE-ε2 dosage 32

eTable 7. Association of clinical risk factors with incident dementia in FGR models in 3C-Dijon. We removed 26 participants with missing information for any of the risk factors. Models were adjusted for age at baseline (incl. squared term), sex, 10 PCs, and dosage of APOE-ε4 and APOE-ε2. 32

eTable 8. Change in estimate of iPRS-DEM after adjustment for clinical risk factors in 3C-Dijon. 26 individuals with missing risk factor data were removed. All models are adjusted for baseline covariates (age at baseline, sex, 10 GPCs) and APOE-ε4 and ε2 dosage. 33

eTable 9. Association of iPRS-DEM with cumulative incidence of dementia censoring at <80 years and from 80 years onwards (>80 years) in 3C-Dijon. models are adjusted for baseline covariates (age at baseline, sex, 10 GPCs) and APOE-ε4 and ε2 dosage. 33

eTable 10a. Comparison of estimated cumulative incidence within APOE4- at different ages in 3C-Dijon . The P value is in reference to the difference between high vs. low iPRS-DEM within APOE4 strata at each age. 34

eTable 10b. Comparison of estimated cumulative incidence within APOE4+ at different ages in 3C-Dijon. The P value is in reference to the difference between high vs. low iPRS-DEM within APOE4 strata at each age..... 34

eTable 10c. Comparison of estimated cumulative incidence across APOE at different ages in 3C-Dijon. Here we test differences in cumulative incidence at different ages across APOE4 strata which allows us to determine how much the risk is driven by APOE4 compared to iPRS-DEM..... 35

eTable 11. Association of APOE, and iPRS-DEM in Memento using Fine-Gray and Cause-Specific hazard models to account for the competing risk of death. All models are adjusted for age at baseline (incl. squared term), sex, 10 PCs. Each model with PRS is also adjusted for APOE-ε4 and APOE-ε2 dosage. 36

eTable 12. Association of clinical risk factors with risk of dementia in Memento. We removed 80 participants with missing information for any of the risk factors. Models were adjusted for age at baseline (incl. squared term), sex, 10 PCs, and dosage of APOE-ε4 and APOE-ε2. 36

eTable 13. Change in estimate of iPRS-DEM after adjustment for clinical risk factors in Memento. We removed 80 participants with missing information for any of the risk factors. All models are adjusted for baseline covariates (age at baseline, sex, 10 GPCs) and APOE-ε4 and ε2 dosage. 36

eTable 14. Association of iPRS-DEM with cumulative incidence of dementia censoring at <80 years and from 80 years onwards (>80 years) in Memento. Models are adjusted for baseline covariates (age at baseline, sex, 10 GPCs) and APOE-ε4 and ε2 dosage. 37

eTable 15a. Comparison of estimated cumulative incidence within APOE4- at different times since baseline assessment in Memento. The P value is in reference to the difference between high vs. low iPRS-DEM within APOE4 strata at each age 37

eTable 15b. Comparison of estimated cumulative incidence within APOE4+ at different times since baseline assessment in Memento. The P value is in reference to the difference between high vs. low iPRS-DEM within APOE4 strata at each age. 38

eTable 15c. Comparison of estimated cumulative incidence across APOE at different times since baseline assessment in Memento. Here we test differences in cumulative incidence at different follow-up times across APOE4 strata which allows us to determine how much the risk is driven by APOE4 compared to iPRS-DEM.. 39

eTable 16. Association of iPRS and with dementia within ancestry sub-groups in All of Us using Cox proportional hazard models. Adjusted for baseline covariates (age, sex, 5PCs) and APOE dosage. EAS = East Asian; AFR = African; HIS = Hispanic; EUR = European 40

eTable 1. *ICD-9 and ICD-10 codes to identify dementia cases in All of Us.*

Dementia		
ICD10	F02	Dementia in other diseases classified elsewhere
ICD10	F03.9	Unspecified dementia
ICD10	I67.3	Progressive vascular leukoencephalopathy
ICD10	F02.81	Dementia in other diseases classified elsewhere with behavioral disturbance
ICD10	G30.0	Alzheimer's disease with early onset
ICD10	G30.8	Other Alzheimer's disease
ICD10	G31.89	Other specified degenerative diseases of nervous system
ICD10	G31.09	Other frontotemporal dementia
ICD10	F02.80	Dementia in other diseases classified elsewhere without behavioral disturbance
ICD10	F01.50	Vascular dementia without behavioral disturbance
ICD10	G30.9	Alzheimer's disease, unspecified
ICD10	G30.1	Alzheimer's disease with late onset
ICD10	F03	Unspecified dementia
ICD10	F03.91	Unspecified dementia with behavioral disturbance
ICD10	G31.0	Frontotemporal dementia
ICD10	G31.1	Senile degeneration of brain, not elsewhere classified
ICD10	G31.83	Dementia with Lewy bodies
ICD10	G31.85	Corticobasal degeneration
ICD10	F03.90	Unspecified dementia without behavioral disturbance
ICD10	G31.84	Mild cognitive impairment, so stated
ICD10	G31.01	Pick's disease
ICD10	G30	Alzheimer's disease

ICD10	F01.51	Vascular dementia with behavioral disturbance
ICD9	331	Alzheimer's disease
ICD9	294.1	Dementia in conditions classified elsewhere
ICD9	291.2	Alcohol-induced persisting dementia
ICD9	290.21	Senile dementia with depressive features
ICD9	331.1	Frontotemporal dementia
ICD9	290.4	Vascular dementia
ICD9	290.41	Vascular dementia, with delirium
ICD9	294.1	Dementia in conditions classified elsewhere without behavioral disturbance
ICD9	290.2	Senile dementia with delusional features
ICD9	290.4	Vascular dementia, uncomplicated
ICD9	290.43	Vascular dementia, with depressed mood
ICD9	290.42	Vascular dementia, with delusions
ICD9	331.11	Pick's disease
ICD9	331.2	Senile degeneration of brain

eTable 2. *Input traits into iPRS-DEM model and associated GWAS references.*

Phenotype†	Abbreviation	GWAS	Citations
Alzheimer's Disease*	AD	Bellenguez et al. 2022 (EADI removed; Memento removed from EADB); *APOE* region removed	Bellenguez C, Küçükali F, Jansen IE, et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. Nat Genet. 2022;54(4):412-436. doi:10.1038/s41588-022-01024-z
All-Stroke*	All_Stroke_EUR All_Stroke_TRANS	Mishra et al. 2022	Mishra A, Malik R, Hachiya T, et al. Stroke genetics informs drug discovery and risk prediction across ancestries. *Nature*. 2022;611(7934):115-123. doi:10.1038/s41586-022-05165-3
Ischemic Stroke*	Ischemic_Stroke_EUR Ischemic_Stroke_TRANS	Mishra et al. 2022	Mishra A, Malik R, Hachiya T, et al. Stroke genetics informs drug discovery and risk prediction across ancestries. *Nature*. 2022;611(7934):115-123. doi:10.1038/s41586-022-05165-4
Small Vessel Stroke *	SV_Stroke_EUR SV_Stroke_TRANS	Mishra et al. 2022	Mishra A, Malik R, Hachiya T, et al. Stroke genetics informs drug discovery and risk prediction across ancestries. *Nature*. 2022;611(7934):115-123. doi:10.1038/s41586-022-05165-5
Cardioembolic stroke*	CES_Stroke_EUR CES_Stroke_TRANS	Mishra et al. 2022	Mishra A, Malik R, Hachiya T, et al. Stroke genetics informs drug discovery and risk prediction across ancestries. *Nature*. 2022;611(7934):115-123. doi:10.1038/s41586-022-05165-6
White matter hyperintensities	WMH	Sargurupremraj et al. 2020	Sargurupremraj M, Suzuki H, Jian X, et al. Cerebral small vessel disease genomics and its implications across the lifespan. *Nat Commun*. 2020;11(1):6285. doi:10.1038/s41467-020-19111-2
Perivascular spaces, basal ganglia, European	PVS_BG	Duperron et al. 2023	Duperron MG, Knol MJ, Le Grand Q, et al. Genomics of perivascular space burden unravels early mechanisms of cerebral small vessel disease. Nat Med. 2023;29(4):950-962. doi:10.1038/s41591-023-02268-w
Perivascular spaces, Hippocampus, European	PVS_HIP	Duperron et al. 2023	Duperron MG, Knol MJ, Le Grand Q, et al. Genomics of perivascular space burden unravels early mechanisms of cerebral small vessel disease. Nat Med. 2023;29(4):950-962. doi:10.1038/s41591-023-02268-w
Perivascular spaces, white matter, European	PVS_WM	Duperron et al. 2023	Duperron MG, Knol MJ, Le Grand Q, et al. Genomics of perivascular space burden unravels early mechanisms of cerebral small vessel disease. Nat Med. 2023;29(4):950-962. doi:10.1038/s41591-023-02268-w
Systolic Blood Pressure	SBP	Evangelou et al. Nat Genet 2018	Evangelou E, Warren HR, Mosen-Ansorena D, et al. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. Nat Genet. 2018;50(10):1412-1425. doi:10.1038/s41588-018-0205-x
Diastolic Blood Pressure	DBP	Evangelou et al. Nat Genet 2018	Evangelou E, Warren HR, Mosen-Ansorena D, et al. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. Nat Genet. 2018;50(10):1412-1425. doi:10.1038/s41588-018-0205-x
Total cholesterol	TC	Graham et al. 2021	Graham SE, Clarke SL, Wu KHH, et al. The power of genetic diversity in genome-wide association studies of lipids. *Nature*. 2021;600(7890):675-679. doi:10.1038/s41586-021-04064-3
Triglyceride	TG	Graham et al. 2021	Graham SE, Clarke SL, Wu KHH, et al. The power of genetic diversity in genome-wide association studies of lipids. *Nature*. 2021;600(7890):675-679. doi:10.1038/s41586-021-04064-4
LDL-Cholesterol	LDL	Graham et al. 2021	Graham SE, Clarke SL, Wu KHH, et al. The power of genetic diversity in genome-wide association studies of lipids. *Nature*. 2021;600(7890):675-679. doi:10.1038/s41586-021-04064-5
HDL-Cholesterol	HDL	Graham et al. 2021	Graham SE, Clarke SL, Wu KHH, et al. The power of genetic diversity in genome-wide association studies of lipids. *Nature*. 2021;600(7890):675-679. doi:10.1038/s41586-021-04064-6
Coronary Artery Disease	CAD	Nelson et al. 2017	EPIC-CVD Consortium, CARDIoGRAMplusC4D, The UK Biobank CardioMetabolic Consortium CHD working group, et al. Association analyses based on false discovery rate implicate new loci for coronary artery disease. Nat Genet. 2017;49(9):1385-1391. doi:10.1038/ng.3913
Atrial Fibrillation	AF	Nielsen et al. Nat Genet 2018	Nielsen JB, Thorolfsdottir RB, Fritsche LG, et al. Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. Nat Genet. 2018;50(9):1234-1239. doi:10.1038/s41588-018-0171-3
Smoking	SMOKING	Liu et al. 2019	23andMe Research Team, HUNT All-In Psychiatry, Liu M, et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. Nat Genet. 2019;51(2):237-244. doi:10.1038/s41588-018-0307-5
Diabetes	T2D	Mahajan et al. Nat Genet 2018	Mahajan A, Taliun D, Thurner M, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nat Genet. 2018;50(11):1505-1513. doi:10.1038/s41588-018-0241-6
Body Mass Index	BMI	Pulit et al. Hum Mol Genet 2019	Pulit SL, Stoneman C, Morris AP, et al. Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum Mol Genet*. 2019;28(1):166-174. doi:10.1093/hmg/ddy327
CSF phosphorylated TAU	pTAU_CSF	Jansen et al. 2022	Jansen IE, van der Lee SJ, Gomez-Fonseca D, et al. Genome-wide meta-analysis for Alzheimer's disease cerebrospinal fluid biomarkers. Acta Neuropathol (Berl). 2022;144(5):821-842. doi:10.1007/s00401-022-02454-z
CSF Amyloid-Beta	AB_CSF	Jansen et al. 2022	Jansen IE, van der Lee SJ, Gomez-Fonseca D, et al. Genome-wide meta-analysis for Alzheimer's disease cerebrospinal fluid biomarkers. Acta Neuropathol (Berl). 2022;144(5):821-842. doi:10.1007/s00401-022-02454-z
Hippocampal Volume	HIP_Vol	Hibar et al. 2017	Hibar DP, Adams HHH, Jahanshad N, et al. Novel genetic loci associated with hippocampal volume. *Nat Commun*. 2017;8(1):13624. doi:10.1038/ncomms13624
*3C was removed from Stroke GWAS meta-anlayses *3C and Memento removed from AD GWAS meta-analyses			**Additional References:** Chauhan G, Adams HHH, Satizabal CL, et al. Genetic and lifestyle risk factors for MRI-defined brain infarcts in a population-based setting. Neurology. 2019;92(5):e486-e503. doi:10.1212/WNL.0000000000006851 Knol MJ, Lu D, Traylor M, et al. Association of common genetic variants with brain microbleeds. Neurology. 2020;95(24):e3331-e3343. doi:10.1212/WNL.0000000000010852 Brouwer RM, Klein M, Grasby KL, et al. Genetic variants associated with longitudinal changes in brain structure across the lifespan. Nat Neurosci. 2022;25(4):421-432. doi:10.1038/s41593-022-01042-4
†GWAS not included due to insignificant heritability: Large Artery Stroke (Mishra et al. 2022); Brain Infarcts and Silent Brain Infarcts (Chauhan et al. 2019); Cerebral Microbleeds (Knol et al. 2020); Longitudinal hippocampal change (Brouwer et al. 2022); Age-related hippocampal change (Brouwer et al. 2022)			

eTable 3. *Best PRS for each individual trait in 3C-Bordeaux-Montpellier training set. Best model was selected on basis of lowest integrated brier score. Fine-Gray models are adjusted for age at baseline, age-squared , study centre, sex, the first 10 PCs, and dosage of APOE-ε4 and APOE-ε2.*

Trait	PRS Method	Parameters	Number of SNPs	sHR	CI_LOW	CI_HIGH	P	AUC-10	IPA-10	IBS (Selection criteria)	Improvement over APOE*
AF	PRSCS	phi=auto	1086590	0.954	0.8431	1.0785	0.45	0.7171	0.0771	0.033091	N
CES_Stroke_EUR	C+T	r2=0.8;pt=0.01	22880	1.04	0.926	1.1687	0.51	0.7173	0.0767	0.033091	N
HIP_Vol	LDPRED2	p=1e-05;h2=0.292	1181711	1.06	0.9454	1.1879	0.32	0.7159	0.0757	0.033087	N
AB_CSF	C+T	r2=0.01;pt=1e-06	8	1.117	0.9763	1.2778	0.11	0.7181	0.0773	0.033036	Y
AD	LDPRED2	p=8.1e-05;h2=0.069 SPARSE	1143554	1.233	1.0966	1.3854	4.00E-04	0.7305	0.0841	0.032837	Y
All_Stroke_EUR	LDPRED2	p=0.00066;h2=0.045 SPARSE	1182943	1.067	0.9572	1.1891	0.24	0.7173	0.0777	0.033046	Y
All_Stroke_TRANS	LDPRED2	p=0.0019;h2=0.027 SPARSE	1117814	1.067	0.958	1.1879	0.24	0.7173	0.0777	0.033021	Y
BMI	LDPRED2	p=0.0053;h2=0.371 SPARSE	1183169	1.154	1.0268	1.2972	0.016	0.7213	0.0825	0.033014	Y
CAD	LDPRED2	p=0.00066;h2=0.067 SPARSE	1169801	0.921	0.825	1.0272	0.14	0.7187	0.0796	0.033022	Y
CES_Stroke_TRANS	C+T	r2=0.1;pt=0.01	13512	1.069	0.9509	1.2026	0.26	0.7172	0.0774	0.033051	Y
DBP	LDPRED2	p=1e-05;h2=0.18	1171464	0.857	0.7651	0.9605	0.0079	0.7229	0.0786	0.033032	Y
HDL	PRSCS	phi=1e-0	1085020	0.869	0.7739	0.9748	0.017	0.7181	0.0813	0.032896	Y
Ischemic_Stroke_EUR	LDPRED2	p=0.0019;h2=0.074	1182818	1.059	0.9465	1.1844	0.32	0.7175	0.0779	0.033058	Y
Ischemic_Stroke_TRANS	C+T	r2=0.01;pt=1e-05	197	1.1	0.9756	1.2411	0.12	0.7192	0.0779	0.033061	Y
LDL	C+T	r2=0.8;pt=0.001	32386	0.937	0.8371	1.0484	0.26	0.7176	0.0776	0.033085	Y
logTG	PRSCS	phi=1e-0	1082094	1.102	0.9853	1.2333	0.089	0.7181	0.0779	0.033036	Y
pTAU_CSF	C+T	r2=0.01;pt=0.4	63832	1.093	0.9798	1.2195	0.11	0.7176	0.0783	0.033005	Y
PVS_BG	LDPRED2	p=0.00066;h2=0.104	1176361	0.933	0.8311	1.0474	0.24	0.718	0.0771	0.033037	Y
PVS_HIP	C+T	r2=0.2;pt=1e-05	17	0.955	0.8484	1.0757	0.45	0.7163	0.0768	0.033085	Y
PVS_WM	C+T	r2=0.1;pt=0.001	1044	0.911	0.8168	1.016	0.094	0.7173	0.0775	0.033037	Y
SBP	LDPRED2	p=2.8e-05;h2=0.183	1169404	0.943	0.8424	1.0558	0.31	0.7176	0.0769	0.033075	Y
SMOKING	LDPRED2	p=0.00066;h2=0.097	1175457	1.1384	1.0171	1.2743	0.024	0.7179	0.0787	0.032923	Y
SV_Stroke_EUR	C+T	r2=0.01;pt=0.05	25285	1.218	1.0848	1.3672	8.00E-04	0.7244	0.0832	0.032945	Y
SV_Stroke_TRANS	C+T	r2=0.01;pt=0.05	26662	1.251	1.111	1.4094	2.00E-04	0.7274	0.0817	0.033009	Y
T2D	LDPRED2	p=1;h2=0.145 SPARSE	1182028	1.08	0.9595	1.2157	0.2	0.7179	0.0792	0.033069	Y
TC	C+T	r2=0.5;pt=0.0001	14035	0.935	0.8346	1.048	0.25	0.718	0.078	0.033084	Y
WMH	C+T	r2=0.01;pt=0.05	15031	0.95	0.8472	1.0651	0.38	0.7173	0.0778	0.033027	Y

eTable 4. *Single-trait PRS Training weights derived from elastic net. Elastic-net was adjusted for age at baseline, age-squared , study centre, sex, the first 10 PCs, and dosage of APOE-ε4 and APOE-ε2. The PRS weight is derived by dividing the elastic-net coefficient by the empirical standard deviation. Optimal elastic-net parameters: α=0.0787; λ=0.001163819*

Trait	Abbreviation	Number of SNPs	Elastic net coefficient (γ)	Empirical Standard Deviation (σ)	PRS Weight
Alzheimer's Disease	AD	1143554	0.071	0.328	0.217
Body Mass Index	BMI	1183169	0.038	0.394	0.097
Coronary Artery Disease	CAD	1169801	-0.010	0.833	-0.012
Diastolic Blood Pressure	DBP	1171464	-0.047	2.508	-0.019
HDL Cholesterol	HDL	1085020	-0.041	0.533	-0.077
Ischemic Stroke (Trans-ancestry)	Ischemic_Stroke_TRANS	197	0.017	0.402	0.042
log Trigylcerides	logTG	1082094	0.017	0.527	0.032
CSF phosphorylated Tau	pTAU_CSF	63832	0.010	6.399	0.001
Perivascular Spaces White Matter	PVS_WM	1044	-0.014	0.328	-0.042
Smoking	SMOKING	1175457	0.031	0.305	0.102
Small Vessel Stroke (EUR)	SV_Stroke_EUR	25285	0.054	7.964	0.007
Small Vessel Stroke (Trans-Ancestry)	SV_Stroke_TRANS	26662	0.074	7.384	0.010

Note: The following traits were included but had a elastic-net coefficient of zero: CSF Amyloid-β, Type-2 Diabetes, Cardioembolic Stroke (Trans-Ancestry), All Stroke (European), All Stroke (Trans-Ancestry), Ischemic Stroke (EUR), Perivascular Spaces Hippocampus, White matter hyperintensities (WMH), Systollic Blood Pressure, LDL Cholesterol, Total Cholesterol, Perivascular spaces basal ganglia

eTable 5. *Top 250 SNPs ordered by the absolute value of their iPRS-weight calculated across all single-trait PRS included in the integrative model.*

TABLE of 250 SNPs & weights, see supplementary table online, omitted here to save space

eTable 6. *Association of APOE, iPRS and AD-PRS with dementia or death in 3C-Dijon using Fine-Gray and Cox proportional hazard models. All models are adjusted for age at baseline (with a quadratic term), sex, 10 PCs. Each model with PRS is also adjusted for APOE-ε4 and APOE-ε2 dosage*

	Fine-Gray		Cause-Specific Dementia		Cause-Specific Death	
	sHR (95% CI)	P	HR (95% CI)	P	HR (95% CI)	P
APOE-ε4	1.514 (1.208-1.90)	**0.0003**	1.582 (1.272-1.969)	**4.00E-05**	1.202 (0.962-1.501)	0.106
APOE-ε2	0.761 (0.556-1.041)	0.088	0.736 (0.541-1.00)	0.050	0.922 (0.721-1.178)	0.514
iPRS per 1SD	1.152 (1.034-1.283)	**0.01**	1.155 (1.042-1.281)	**0.006**	1.048 (0.952-1.153)	0.341
AD-PRS per 1SD	1.107 (0.996-1.232	0.06	1.10 (1.00-1.231)	**0.049**	0.981 (0.891-1.079)	0.688
iPRS Q1	1.084 (0.768-1.53)	0.65	1.084 (0.773-1.519)	0.642	0.862 (0.633-1.174)	0.346
iPRS Q2	0.939 (0.663-1.329)	0.72	0.947 (0.67-1.34)	0.760	1.023 (0.763-1.371)	0.881
iPRS Q4	1.144 (0.814-1.609)	0.44	1.172 (0.835-1.645)	0.359	1.24 (0.931-1.653)	0.142
iPRS Q5	1.484 (1.076-2.047)	**0.016**	1.479 (1.076-2.033)	**0.016**	0.935 (0.688-1.271)	0.668
AD-PRS Q1	0.9402 (0.667-1.32)	0.72	0.921 (0.655-1.294)	0.634	1.006 (0.744-1.36)	0.970
AD-PRS Q2	0.91 (0.644-1.284)	0.59	0.901 (0.642-1.264	0.545	0.944 (0.70-1.274)	0.707
AD-PRS Q4	1.187 (0.86-1.639)	0.3	1.192 (0.862-1.647)	0.288	1.095 (0.812-1.478)	0.551
AD-PRS Q5	1.211 (0.876-1.674)	0.25	1.19 (0.86-1.641)	0.296	0.915 (0.672-1.247)	0.575

eTable 7. *Association of clinical risk factors with incident dementia in FGR models in 3C-Dijon. We removed 26 participants with missing information for any of the risk factors. Models were adjusted for age at baseline (incl. squared term), sex, 10 PCs, and dosage of APOE-ε4 and APOE-ε2.*

Risk Factor	**Definition**	**sHR per 1 SD in 3C-Dijon**	**CI.LOW**	**CI.HIGH**	**P**
Low Education	Did not complete beyond primary education	1.241	1.129	1.363	**1.00E-05**
Smoking	Current smoker	0.888	0.764	1.032	0.12
Obesity	Body Mass Index ≥ 30 kg/m²	1.179	1.072	1.296	**0.00066**
Hypertension	Blood pressure ≥ 140/90 mmHg or antihypertensive treatment	0.979	0.875	1.097	0.72
Diabetes	Fasting glycemia ≥ 7 mmol/L or taking diabetes medication	1.14	1.042	1.247	**0.0044**
History of CVD	History of: myocardial infarction, hospitalized stroke, coronary surgery/angioplasty, history of leg artery surgery if arteritis of the lower limbs	1.22	1.119	1.33	**1.00E-05**
Dyslipidemia	Fasting blood cholesterol ≥ 6.2 mmol/L or taking lipid-lowering medication	1.022	0.917	1.139	0.69

eTable 8. *Change in estimate of iPRS-DEM after adjustment for clinical risk factors in 3C-Dijon. 26 individuals with missing risk factor data were removed. All models are adjusted for baseline covariates (age at baseline, sex, 10 GPCs) and APOE-ε4 and ε2 dosage.*

Model	sHR of iPRS	CI.LOW	CI.HIGH	Change in Estimate %*
iPRS + Vascular RF + LOW_EDU	1.109	0.994	1.238	27.02
iPRS + Vascular RF	1.123	1.006	1.255	18.15
iPRS + hypct024	1.156	1.037	1.289	-1.91
iPRS + cardvas0	1.135	1.016	1.267	11.05
iPRS + diabete	1.144	1.027	1.276	5.08
iPRS + hta0_1	1.153	1.034	1.286	0.01
iPRS + OBESE	1.142	1.024	1.274	6.42
iPRS + SMOKE	1.152	1.033	1.285	0.34
iPRS + LOW_EDU	1.137	1.02	1.268	9.37
iPRS	1.153	1.034	1.285	0

*Change in estimate is based on relative change in iPRS-DEM coefficient compared to the model only adjusted for age, sex, 10PCs and APOE-ε4 and ε2 dosage.

eTable 9. *Association of iPRS-DEM with cumulative incidence of dementia censoring at <80 years and from 80 years onwards (>80 years) in 3C-Dijon. models are adjusted for baseline covariates (age at baseline, sex, 10 GPCs) and APOE-ε4 and ε2 dosage.*

	Censoring	**N Cases**	**N**	**sHR of iPRS-DEM**	**CI_LOW**	**CI_HIGH**	**P**
3C-Dijon	<80 years	123	3201	1.17	0.96	1.43	0.12
	>80 years	238	2165	1.16	1.01	1.32	0.029
	Men	141	1414	1.04	0.88	1.23	0.66
	Women	220	2288	1.23	1.07	1.41	0.004

eTable 10a. *Comparison of estimated cumulative incidence within APOE4- at different ages in 3C-Dijon . The P value is in reference to the difference between high vs. low iPRS-DEM within APOE4 strata at each age.*

APOE	iPRS-DEM*	Age	Estimated cumulative incidence (%)	lower CI	upper CI	Number at Risk	P.inAPOE
APOE4-	High	75	3.51	-1.42	8.44	81	0.4706
APOE4-	Low	75	1.67	0.94	2.40	1359	ref.
APOE4-	High	80	13.21	5.98	20.45	89	**0.0264**
APOE4-	Low	80	4.92	3.78	6.06	1247	ref.
APOE4-	High	85	21.07	12.40	29.73	46	**0.0345**
APOE4-	Low	85	11.52	9.72	13.32	601	ref.
APOE4-	High	90	37.29	25.94	48.65	13	**0.0057**
APOE4-	Low	90	20.79	17.97	23.61	199	ref.

* iPRS-DEM is specified as high or low using the 93.8th percentile cutoff.

eTable 10b. *Comparison of estimated cumulative incidence within APOE4+ at different ages in 3C-Dijon. The P value is in reference to the difference between high vs. low iPRS-DEM within APOE4 strata at each age.*

APOE	iPRS-DEM*	Age	Estimated cumulative incidence (%)	lower CI	upper CI	Number at Risk	P.inAPOE
APOE4+	High	75	4.26	0.87	7.64	165	0.9303
APOE4+	Low	75	4.06	1.10	7.02	211	ref.
APOE4+	High	80	12.38	7.37	17.39	135	0.6138
APOE4+	Low	80	10.67	6.33	15.01	160	ref.
APOE4+	High	85	28.54	20.97	36.10	53	**0.0060**
APOE4+	Low	85	15.51	10.10	20.92	74	ref.
APOE4+	High	90	40.06	30.68	49.43	10	**0.0134**
APOE4+	Low	90	24.18	15.77	32.58	17	ref.

* iPRS-DEM is specified as high or low using the 55.6th percentile cutoff.

eTable 10c. *Comparison of estimated cumulative incidence across APOE at different ages in 3C-Dijon. Here we test differences in cumulative incidence at different ages across APOE4 strata which allows us to determine how much the risk is driven by APOE4 compared to iPRS-DEM*

Comparison	Age	P
APOE4- & Low PRS vs. APOE4+ & Low PRS	75	0.1254
APOE4- & Low PRS vs. APOE4+ & Low PRS	80	**0.012**
APOE4- & Low PRS vs. APOE4+ & Low PRS	85	0.1703
APOE4- & Low PRS vs. APOE4+ & Low PRS	90	0.4543
APOE4- & Low PRS vs. APOE4+ & High PRS	75	0.1434
APOE4- & Low PRS vs. APOE4+ & High PRS	80	**0.0044**
APOE4- & Low PRS vs. APOE4+ & High PRS	85	**<1x10e-4**
APOE4- & Low PRS vs. APOE4+ & High PRS	90	**1.00E-04**
APOE4- & High PRS vs. APOE4+ & Low PRS	75	0.8518
APOE4- & High PRS vs. APOE4+ & Low PRS	80	0.5551
APOE4- & High PRS vs. APOE4+ & Low PRS	85	0.2862
APOE4- & High PRS vs. APOE4+ & Low PRS	90	0.0687
APOE4- & High PRS vs. APOE4+ & High PRS	75	0.8061
APOE4- & High PRS vs. APOE4+ & High PRS	80	0.8527
APOE4- & High PRS vs. APOE4+ & High PRS	85	0.203
APOE4- & High PRS vs. APOE4+ & High PRS	90	0.7131

eTable 11. *Association of APOE, and iPRS-DEM in Memento using Fine-Gray and Cause-Specific hazard models to account for the competing risk of death. All models are adjusted for age at baseline (incl. squared term), sex, 10 PCs. Each model with PRS is also adjusted for APOE-ε4 and APOE-ε2 dosage.*

	Fine-Gray		**Cause-Specific: Dementia**		**Cause-Specific: Death**	
	sHR (95% CI)	P	HR (95% CI)	P	HR (95% CI)	P
APOE-ε4	2.478 (2.069-2.967)	**<0.0001**	2.491 (2.082-2.981)	**<0.0001**	1.152 (0.656-2.022)	0.623
APOE-ε2	0.535 (0.325-0.88)	0.014	0.527 (0.328-0.849)	**0.0084**	0.612 (0.238-1.576)	0.309
iPRS per 1SD	1.254 (1.111-1.416)	**0.0003**	1.256 (1.117-1.413)	**0.0001**	0.802 (0.601-1.07)	0.133
iPRS Q1	0.732 (0.484-1.107)	0.14	0.73 (0.488-1.091)	0.125	3.068 (1.111-8.473)	**0.031**
iPRS Q2	0.996 (0.67-1.48)	0.98	0.988 (0.671-1.456)	0.952	2.875 (0.997-8.293)	0.051
iPRS Q4	1.071 (0.734-1.563)	0.72	1.059 (0.726-1.544)	0.767	1.134 (0.342-3.756)	0.838
iPRS Q5	1.58 (1.101-2.268)	**0.013**	1.579 (1.104-2.26)	**0.012**	1.986 (0.65-6.068)	0.228

eTable 12. *Association of clinical risk factors with risk of dementia in Memento. We removed 80 participants with missing information for any of the risk factors. Models were adjusted for age at baseline (incl. squared term), sex, 10 PCs, and dosage of APOE-ε4 and APOE-ε2.*

Risk Factor	**Definition**	**sHR per 1 SD**	**CI.LOW**	**CI.HIGH**	**P**
Low Education	Did not complete beyond primary education	1.12	1.012	1.241	**0.029**
Family History of Dementia	2nd degree relative diagnosed with dementia: grandparent, mother, father, child, brother, sister, uncle, aunt or first cousin	1.134	1.009	1.275	**0.035**
Hypertension	Blood pressure ≥ 140/90 mmHg or antihypertensive treatment	0.879	0.779	0.992	**0.037**
Smoking	Current Smoker	0.997	0.867	1.147	0.96
Diabetes	diagnosed diabetes at baseline assessment	1.053	0.943	1.176	0.36
Dyslipidemia	dyslipidemia treated at inclusion	1.064	0.949	1.193	0.29
History of Cardiovascular Disease	anyone who recordered coronary artery disease, stroke, myocardial infarction, or cardiac valve issues in their medical history	1.014	0.905	1.135	0.82
Obesity	Body Mass Index ≥ 30 kg/m²	0.918	0.803	1.049	0.21

*Risk factors are all binary, but standardized in order to allow for comparison with iPRS per 1SD increase.

eTable 13. *Change in estimate of iPRS-DEM after adjustment for clinical risk factors in Memento. We removed 80 participants with missing information for any of the risk factors. All models are adjusted for baseline covariates (age at baseline, sex, 10 GPCs) and APOE-ε4 and ε2 dosage.*

model	HR	CI.LOW	CI.HIGH	Change in Estimate %
iPRS + All RF	1.243	1.101	1.404	4.42
iPRS + fmh_dem	1.254	1.111	1.417	0.56
iPRS + lowEDU	1.246	1.104	1.406	3.49
iPRS	1.256	1.112	1.418	0

*Change in estimate is based on relative change in iPRS-DEM beta compared to model only adjusted for age, sex, 10PCs and APOE-ε4 and ε2 dosage

eTable 14. *Association of iPRS-DEM with cumulative incidence of dementia censoring at <80 years and from 80 years onwards (>80 years) in Memento. Models are adjusted for baseline covariates (age at baseline, sex, 10 GPCs) and APOE-ε4 and ε2 dosage.*

	Censoring	N Cases	N	HR of iPRS-DEM	CI_LOW	CI_HIGH	P
Memento	<80 years	172	1758	1.3	1.11	1.52	9E-04
	>80 years	112	638	1.13	0.92	1.38	0.25
	Men	133	782	1.17	0.97	1.4	0.11
	Women	151	1250	1.3	1.1	1.54	0.002

eTable 15a. *Comparison of estimated cumulative incidence within APOE4- at different times since baseline assessment in Memento. The P value is in reference to the difference between high vs. low iPRS-DEM within APOE4 strata at each age*

APOE	iPRS-DEM*	Years since baseline	Estimated cumulative incidence (%)	lower CI	upper CI	Number at Risk	P.inAPOE
APOE4-	High	1	3.26	1.40	5.12	335	0.106
APOE4-	Low	1	1.60	0.85	2.36	1007	Ref.
APOE4-	High	2	5.82	3.29	8.36	300	0.109
APOE4-	Low	2	3.55	2.42	4.69	931	Ref.
APOE4-	High	3	8.53	5.43	11.63	273	0.218
APOE4-	Low	3	6.36	4.83	7.88	843	Ref.
APOE4-	High	4	13.12	9.19	17.04	240	**0.016**
APOE4-	Low	4	7.84	6.14	9.54	785	Ref.
APOE4-	High	5	13.78	9.75	17.82	204	0.078
APOE4-	Low	5	9.77	7.85	11.70	675	Ref.

* iPRS-DEM is specified as high or low using the 74.1th percentile cutoff.

eTable 15b. *Comparison of estimated cumulative incidence within APOE4+ at different times since baseline assessment in Memento. The P value is in reference to the difference between high vs. low iPRS-DEM within APOE4 strata at each age.*

APOE	iPRS-DEM*	Years since baseline	Estimated cumulative incidence (%)	lower CI	upper CI	Number at Risk	P.inAPOE
APOE4+	High	1	5.48	1.88	9.08	158	0.933
APOE4+	Low	1	5.67	3.24	8.10	357	Ref.
APOE4+	High	2	17.53	11.29	23.78	134	0.085
APOE4+	Low	2	11.35	8.11	14.58	317	Ref.
APOE4+	High	3	30.12	22.53	37.71	105	**0.002**
APOE4+	Low	3	16.62	12.79	20.45	284	Ref.
APOE4+	High	4	35.00	26.96	43.04	96	**0.001**
APOE4+	Low	4	19.67	15.57	23.76	248	Ref.
APOE4+	High	5	41.30	32.86	49.73	72	**7.20E-05**
APOE4+	Low	5	22.12	17.82	26.42	215	Ref.

eTable 15c. *Comparison of estimated cumulative incidence across APOE at different times since baseline assessment in Memento. Here we test differences in cumulative incidence at different follow-up times across APOE4 strata which allows us to determine how much the risk is driven by APOE4 compared to iPRS-DEM*

Comparison	Time	P
APOE4- & Low PRS vs. APOE4+ & Low PRS	1	0.0018
APOE4- & Low PRS vs. APOE4+ & Low PRS	2	<0.0001
APOE4- & Low PRS vs. APOE4+ & Low PRS	3	<0.0001
APOE4- & Low PRS vs. APOE4+ & Low PRS	4	<0.0001
APOE4- & Low PRS vs. APOE4+ & Low PRS	5	<0.0001
APOE4- & Low PRS vs. APOE4+ & High PRS	1	0.039
APOE4- & Low PRS vs. APOE4+ & High PRS	2	<0.0001
APOE4- & Low PRS vs. APOE4+ & High PRS	3	<0.0001
APOE4- & Low PRS vs. APOE4+ & High PRS	4	<0.0001
APOE4- & Low PRS vs. APOE4+ & High PRS	5	<0.0001
APOE4- & High PRS vs. APOE4+ & Low PRS	1	0.1239
APOE4- & High PRS vs. APOE4+ & Low PRS	2	0.0085
APOE4- & High PRS vs. APOE4+ & Low PRS	3	0.0013
APOE4- & High PRS vs. APOE4+ & Low PRS	4	0.0236
APOE4- & High PRS vs. APOE4+ & Low PRS	5	0.0056
APOE4- & High PRS vs. APOE4+ & High PRS	1	0.284
APOE4- & High PRS vs. APOE4+ & High PRS	2	7.00E-04
APOE4- & High PRS vs. APOE4+ & High PRS	3	<0.0001
APOE4- & High PRS vs. APOE4+ & High PRS	4	<0.0001
APOE4- & High PRS vs. APOE4+ & High PRS	5	<0.0001

eTable 16. *Association of iPRS and with dementia within ancestry sub-groups in All of Us using Cox proportional hazard models. Adjusted for baseline covariates (age, sex, 5PCs) and APOE dosage. EAS = East Asian; AFR = African; HIS = Hispanic; EUR = European*

Ancestry							
EAS		**AFR**		**HIS**		**EUR**	
HR [95% CI]	p-value	HR [95% CI]	p-value	HR [95% CI]	p-value	HR [95% CI]	p-value
5.29 (1.02 - 27.42)	0.047	1.06 (0.76 - 1.47)	0.75	1.09 (0.74 - 1.62)	0.66	1.28 (1.10 - 1.51)	0.003

Appendix 7

Polygenic risk informs anticoagulation treatment in intracerebral hemorrhage survivors with atrial fibrillation

Supplement

Table of Contents

Supplementary Methods 2

Quality control of Genotyped Data for Polygenic Risk Scoring 2

1. Pre-variant Quality Control 2

2. Sample quality control 2

3. Variant quality control 3

4. HRC imputation 3

5. Post-Imputation QC 3

Endpoint and Outcome Definitions in PRESTIGE-AF Trial 4

Index Event Bias Analysis 9

UK Biobank ICH Sample Selection 9

Inverse probability weighting 10

Supplementary Tables 11

Supplementary Methods

Quality control of Genotyped Data for Polygenic Risk Scoring

1. Pre-variant Quality Control

First, in order to exclude bad quality variants from the assessments of samples, we performed a pre- variant QC where we removed variants on both autosomes and chromosome X having a Hardy- Weinberg equilibrium test p-value<0.001, showing a missingness >0.02 and with a Minor Allele Frequency <0.01 (485,870 variants removed).

2. Sample quality control

Heterozygosity and missingness

Using variants passing the pre-variant QC, we assessed each sample by computing the missingness globally on autosomes, we removed samples having a missingness >0.05 (0 sample removed).

We then computed the heterozygosity using Plink and generated the heterozygosity vs missingness plot to determine the threshold to use for exclusions. Based on this plot, we removed participants outside the interval mean +/- 6SD (1 sample removed).

Sexcheck

Using the samples not failing the Missingness/Heterozygosity QC and the variants on the chromosome X passing the pre-variant QC, we performed a sexcheck test using plink. First, we removed variants in the PAR region (using the correct assembly – GRCh37), then we pruned the remaining variants (using a 0.2 r^2 threshold of linkage disequilibrium, distance 500kb) and performed the sexcheck. We chose to exclude:

- Samples with clinical sex of 1 (male) with F < 0.7
- Samples with clinical sex of 2 (female) F < 0.3
- 1 sample was removed at this step.

Relatedness estimation

We used the KING Software to generate the kinship coefficients. We used all the autosomal SNPs (without filtering on the MAF) passing the same check than in the Pre-SNP QC (HWE > 0.001 and missingness < 0.02).

After generating the kinship estimation, we identified and removed the samples potentially contaminated:

- Samples related to more than 5 other participants using a kinship threshold >0.10 (0 sample removed)
- Samples related to more than 20 other participants using a kinship threshold >0.0625 (3rd degree related) (0 sample removed)

We then generated the kinship estimates for all pairs, we extracted pairs with kinship >0.10 and kinship >0.0625: **No related pairs were flagged.**

Ancestry

On the remaining set of 128 samples, we performed an ancestry check against 1000 Genomes Phase3 (EUR sample).

- PRESTIGE-AF genotypes (n=128)
- 1000 Genomes Phase3 data in GRCh37 – European sample only (n=503)

First, we selected a subset of variants to include in the ancestry analysis :

- variants in common between the three datasets.
- minor allele frequency (MAF) > 0.01 in both datasets
- not ambiguous : not A/T or C/G
- not located in high LD regions.

We then pruned the remaining variants (0.2 of R^2 of linkage disequilibrium, distance 500kb) and merged the pruned datasets.

Principal components (PCs) were computed using Plink on the merged dataset. After graphical review, we chose to flag samples as ancestry outliers on PC1 and PC2 outside the interval mean ± 4SD.

Using PC1 and PC2, no ancestry outliers were identified.

3. Variant quality control

From the initial sets of variant (re-integrating variants failing pre-variant QC), excluding samples failing the sample QC (missingness, heterozygosity, sex check or relatedness check), we have excluded the following variants:

- Non-autosomal (n=37,477)
- Non bi-allelic (n=62,413)
- Missingness > 2% (n=10,288)
- Hardy-Weinberg p < 0.001 (n=968)
- 111,146 variants were removed at this step.

Sample QCed before pre-imputation check: 128 samples & 809,529 variants

4. HRC imputation

As a final step before HRC imputation, we used the McCarthy group pipeline for HRC pre-imputation Checks (https://www.well.ox.ac.uk/~wrayner/tools/) as specified in the Helmholtz Munich Imputation Server help page to flip and swap alleles according to the reference panel. We used the Helmholtz Munich Imputation Server.

After imputation, 39,082,989 variants were available in the dataset.

5. Post-Imputation QC

We performed post-imputation quality control to prepare the data for polygenic risk scoring:

- Remove Duplicates

- Imputed SNP has RSID
- MAF <0.01
- INFO > 0.8
- Call Rate > 0.95

Table 1. HRC SNP- Results

	HRC
Sample size	128
All SNPs in file	39,082,989
Duplicate SNP IDs	44,634
Missing RSID	6,245,717
Number of SNPs with MAF == 0	8,544,307
number of SNPs with MAF >= 0.01	7,920,113
number of SNPs with MAF < 0.01 & >=0.001	4,043,946
number of SNPs with MAF < 0.001 & >=0.0001	3,608,268
number of SNPs with MAF < 0.0001 & >=0.00001	5,619,121
SNPs with INFO > 0.8, MAF > 0.01, Call Rate > 0.95	6,980,126

Endpoint and Outcome Definitions in PRESTIGE-AF Trial

Primary Stroke Endpoints	Secondary Endpoints
There are two co-primary binary endpoints in the study: • Ischemic stroke • Recurrent Intracerebral Hemorrhage Both co-primary endpoints were also analyzed together as **any stroke**.	Major bleeding safety endpoints comprise of: • Any major hemorrhage according to International Society of Thrombosis and Hemostasis (ISTH) bleeding assessment tool, including: o Intracranial hemorrhage o Intracerebral hemorrhage Net clinical benefit defined as composite endpoint of all stroke, systemic embolism, myocardial infarction, cardiovascular mortality, and major bleeding.

Outcome event		Definition
Stroke		Stroke is acute focal brain dysfunction due to a vascular cause lasting ≥ 24 hours in the absence of brain imaging or requires evidence of acute stroke on brain imaging (i.e. if there is an acute/subacute stroke documented by CT or MRI or at autopsy, the duration of symptoms/signs may be less than 24 hours). Stroke does not include diffuse brain ischemia from hypoxic-ischemic encephalopathy. If death occurs within 24 hours, the neurological deficit must persist up to the time of death.
	Ischemic stroke	An episode of neurological dysfunction caused by focal cerebral, spinal, or retinal infarction. Evidence of ischemic infarction is defined as brain, spinal cord, or retinal cell death attributable to ischemia, based on pathological, imaging, or other objective evidence of cerebral, spinal cord, or retinal focal ischemic injury in a defined vascular distribution; or clinical evidence of cerebral, spinal cord, or retinal focal ischemic injury based on symptoms persisting ≥24 hours or until death, and other etiologies excluded
	Hemorrhagic stroke	Hemorrhagic stroke (i.e. intracerebral hemorrhage) is defined as rapidly developing clinical signs of neurological dysfunction attributable to a focal collection of blood within the brain parenchyma or ventricular system that is not caused by trauma. Hemorrhagic stroke requires neuroimaging or autopsy confirmation and includes two subcategories: symptomatic primary intracerebral hemorrhage (intraparenchymal or intraventricular) and symptomatic primary subarachnoid hemorrhage. Intracranial bleeding caused by head trauma, bleeding associated with tumours, and subdural/epidural haematomas are not considered as hemorrhagic strokes (but these will be counted separately as major hemorrhages).
	Primary intracerebral hemorrhage	Rapidly developing clinical signs of neurological dysfunction attributable to a focal collection of blood within the brain parenchyma or ventricular system that is not caused by trauma These are symptomatic hemorrhagic strokes with CT/MRI or autopsy evidence of bleeding into the substance of the brain or ventricular spaces. Large or superficial intracerebral hemorrhages often are associated with minor amounts of subarachnoid hemorrhage, but these should be classified as intracerebral hemorrhages. NOTE: Does not include secondary hemorrhage into cerebral infarct (i.e. hemorrhagic transformation), or

		intracerebral bleeding (i.e. contusions) due to trauma, or microbleeds detected by MRI.
	Primary subarachnoid hemorrhage	Typical clinical syndrome of sudden onset headache, with or without focal signs (subarachnoid hemorrhage may not have focal deficits), and CT or cerebrospinal fluid evidence of bleeding primarily into the subarachnoid space. Subarachnoid bleeding due to ruptured intracranial aneurysms and vascular malformation are counted as hemorrhagic strokes, but traumatic subarachnoid hemorrhage is not (but counted separately as major bleeding).
	Undetermined stroke	Definite stroke that does not meet the above criteria for ischemic or hemorrhagic stroke because CT scan or MRI are not done within 3 weeks and there are no autopsy data. A focal neurological deficit must be documented to be present for 24 hours or more. Rarely it cannot be determined with confidence whether the stroke was ischemic vs. hemorrhagic, even after review of CT/MRI images (e.g primary intracerebral hemorrhage vs. severe hemorrhagic transformation); these stroke events will be classified as undetermined.
Major Adverse Cardiac Events (MACE)		• Angiographically confirmed Stent Thrombosis • Acute Coronary Syndrome • Myocardial Ischemia • Myocardial Infarction • Cardiac Arrest • Systemic Embolism • Cardiovascular death
	Myocardial infarction (MI)	The term acute MI is used when there is evidence of myocardial necrosis in a clinical setting consistent with acute myocardial ischemia. The diagnosis of MI requires the combination of evidence of: • Myocardial necrosis (either changes in cardiac biomarkers (preferably cardiac troponin with at least one value above the 99th percentile) or post-mortem pathological findings) and, at least one of the five following diagnostic criteria should be met: • Symptoms of ischemia • New (or presumably new) significant ST/T wave changes or left bundle-branch block (LBBB) • Development of pathological Q waves on ECG • Imaging evidence of new loss of viable myocardium or new regional wall motion abnormality

		• Identification of intracoronary thrombus by angiography or autopsy The '5 Types of Myocardial Infarction' are described below from the MI Universal Definition (2012), • Type 1: Spontaneous MI • Type 2: MI secondary to an ischemic imbalance • Type 3: MI resulting in death when biomarker values are unavailable • Type 4a: MI related to percutaneous coronary intervention (PCI) • Type 4b: MI related to stent thrombosis • Type 5: MI related to coronary artery bypass grafting (CABG)
Systemic arterial embolism		Systemic arterial embolism is defined as abrupt vascular insufficiency associated with evidence of arterial occlusion in the absence of other likely mechanisms. Clinical signs and symptoms must be consistent with embolic arterial occlusion, there must be clear evidence of abrupt occlusion of a systemic artery, with at least one type of supporting evidence, such as surgical report indicating evidence of arterial embolism, pathological specimens related to embolism removal, imaging evidence consistent with arterial embolism, or autopsy report. NOTE: Excluded from the definition are pulmonary embolism and non-embolic arterial occlusions.
Death		All deaths will be primarily classified according to whether they are cardiovascular or not. The mandated approach is to consider a death to be cardiovascular unless there is a clear non- cardiovascular cause. Furthermore, deaths due to hemorrhage will be considered cardiovascular unless the hemorrhage was trauma induced. Deaths that occur unwitnessed will be considered cardiovascular. Deaths due to cardiovascular diagnostic or therapeutic procedures will be considered cardiovascular. Unwitnessed death out of hospital will be considered cardiovascular.
	Cardiovascular death	• Myocardial Infarction • Sudden cardiac death • Death due to heart failure • Death due to stroke • Death due to cardiovascular procedures • Death due to other cardiovascular cause
	Non cardiovascular death	• Cancer • Respiratory failure (unless as a complication of stroke) • Liver failure • Infection / sepsis (unless as a complication of stroke)

		• Renal failure • Accident or trauma due to road traffic accident or other injury (i.e. traumatic intracranial hemorrhage) • Suicide • Other non-vascular cause
	Unknown death	Undetermined or unknown death refers to a death not attributable to one of the above categories of cardiovascular death or to a non-cardiovascular cause. Inability to classify the cause of death may be due to lack of information (e.g., the only available information is "patient died"). These deaths will be considered as unknown cause but will be treated as cardiovascular in the primary analysis of study results.
Major bleeding		An acute or subacute clinically overt bleeding event that meets at least one of the following criteria: • Fatal bleeding, and/or • Symptomatic bleeding in a critical area or organ (intracranial, intraocular, intraspinal, pericardial, retroperitoneal, intraarticular, or intramuscular with compartment syndrome), and/or • Clinically overt bleeding associated with a recent decrease in the haemoglobin level of ≥2 g/dL (20 g/L; 1.24 mmol/L) compared to the most recent haemoglobin value available before the event, and/or • Clinically overt bleeding leading to transfusion of 2 or more units of packed red blood cells or whole blood
Intracranial hemorrhage		Symptomatic intracerebral/intraparenchymal hemorrhages as well as symptomatic subarachnoid hemorrhages will be captured as bleeding and as stroke. All subdural/epidural haematoma and asymptomatic intracranial bleeding are only captured as bleeding. Hemorrhagic transformation of an ischemic stroke will not be counted as intracranial hemorrhage; microbleeds detected are not considered intracranial hemorrhage.

Index Event Bias Analysis

UK Biobank ICH Sample Selection

The first step was to identify ICH survivors with AF in the UK Biobank. Individuals were excluded if they were not of European ancestry. We also removed everyone with self-reported AF or ICH to avoid possible misclassification due to inaccurate self-reporting.

Exclusions: Total UK Biobank N = 502,336

- Non-European ancestry: n = 93,110
- Self-reported ICH: n =118
- Self-reported AF: n = 663

We first used algorithmically defined outcomes from the UK Biobank to identify 2,054 individuals who experienced an ICH (field code 42010).

Exclusions:

- Did not survive past 30 days: n = 640
- Possible hemorrhagic transformation (Ischemic stroke on or up to 30 days before): n = 115

N = 1,299 ICH survivors

Individuals with AF were identified using field code 131350, and self-reported AF cases were also removed. Of the remaining 1,299 ICH cases, 240 had AF documented on or before the date of their ICH.

The control population was defined as the remained of the UK Biobank of European ancestry. To avoid potential misclassification in the control population we removed individuals with possible hemorrhagic transformation or with self-reported stroke as well, which were all of an undefined type. 408,033 controls were used.

Next we defined variables for smoking, congestive heart failure, type-2 diabetes, ischemic stroke history, hypertension, sex, and age at baseline.

Table: UK Biobank variables included in the selection model

Variable	**Field code**	**Coding**
Sex	31	0= Female 1 = Male
Age at Baseline assessment	21003	Continuous

Hypertension	131286 (ICD-10 code I10)	Binary
Congestive heart failure	131354 (ICD-10 code I50)	Binary
Diabetes (insulin-dependent)	130706 (ICD-10 code E10)	Binary
Previous ischemic stroke	42008 (algorithmically-defined outcome)	Binary
Ever Smoker	20116 (self-report	Binary

Our final sample size constituted 237 ICH survivors with AF, and 407,225 controls who had complete covariate information.

Inverse probability weighting

We used inverse probability weighting to adjust for index event bias by reweighting people in the analysis to account for the likelihood o f being selected into the study (i.e., being an ICH survivor with AF)[1,2]. This works by first estimating a selection model in the UK Biobank. Then we re-weight individuals in the target sample (PRESTIGE-AF) by the inverse of their predicted probabilities based on the selection model. These weights adjust associations with individuals with high selection probability being down-weighted, and those with low selection probability are up-weighted.

The selection model had the following weights derived from logistic regression:

Covariate	**Estimate**	**Std. Error**	**z value**	**Pr(>|z|)**
intercept	-16.943	0.875	-19.35	1.86E-83
age at baseline	0.133	0.014	9.65	5.10E-22
sex	0.686	0.143	4.80	1.61E-06
hypertension	1.222	0.176	6.95	3.76E-12
diabetes	0.457	0.364	1.26	2.09E-01
previous IS	0.874	0.200	4.37	1.22E-05
CHF	0.604	0.171	3.53	4.18E-04
Ever Smoker	0.084	0.134	0.63	5.29E-01

Baseline table comparing UK Biobank controls, ICH survivors with AF, and PRESTIGE-AF

	Controls	**ICH survivors with AF (UKBB)**	**PRESTIGE-AF**
N	407,225	237	128
Age at baseline mean (SD)	57 (8.0)	64 (4.8)	77 (8.5)
Male Sex (%)	187,418 (46.0)	163 (68.8)	83 (64.8)
Congestive heart failure	18309 (4.5)	46 (19.4)	20 (15.6)
Diabetes	4,100 (1.0)	8 (3.4)	40 (31.2)
Previous ischemic stroke	9,788 (2.4)	30 (12.7)	25 (19.5)

Hypertension	166,162 (40.8)	195 (82.3)	123 (96.1)
Ever Smoker	184,929 (45.4)	136 (57.4)	60 (46.9)

References

1. Cole SR, Hernán MA. Constructing Inverse Probability Weights for Marginal Structural Models. *Am J Epidemiol*. 2008;168(6):656-664. doi:10.1093/aje/kwn164

2. Mitchell RE, Hartley AE, Walker VM, et al. Strategies to investigate and mitigate collider bias in genetic and Mendelian randomisation studies of disease progression. *PLOS Genetics*. 2023;19(2):e1010596. doi:10.1371/journal.pgen.1010596

Supplementary Tables

Supplementary Table 1. PRS model sources and number of SNPs in PRESTIGE-AF.

PRS	**GWAS/PRS publication**	**Number of SNPs - Publication/GWAS**	**N SNPs in PRESTIGE-AF PRS**
Atrial Fibrillation	PMID: 36459520 Downloaded from: https://cvd.hugeamp.org/downloads.html#polygenic	1,090,061	1,016,157 (93.2%)
Integrative PRS for Ischemic stroke	PMID: 36180795 Catalog: PGS002724	1,213,574	1,189,524 (98%)
Ischemic Stroke GRS	PMID: 36180795	5,493,388 out of 7,464,047 (74.9%) available in PRESTIGE-AF after QC	(r2<0.1 & p<5e-08) 33 SNPs
Intracerebral Hemorrhage	PMID: 36756894 Catalog: PGS003457	682,890	650,860 (95.3%)
White Matter Hyperintensities	PMID: 33293549	4,868,146 out of 6,729,562 available (72.3%)	(r2<0.1 & p<5e-08) 28 SNPs

Supplementary Table 2a. Associations of PRS with primary and secondary endpoints adjusted for CHA2DS2-VASc

	Any Stroke		Ischemic Stroke		ICH		Major Bleed		MACE	
	HR 95% CI	**P**	**HR 95% CI**	**P**	**HR 95% CI**	**P**	**HR 95% CI**	**P**	**HR 95% CI**	**P**
AF-PRS	0.91 [0.54-1.53]	0.72	1.33 [0.66-2.68]	0.42	0.39 [0.15-1.04]	0.06	0.46 [0.21-0.99]	0.05	0.79 [0.5-1.26]	0.33
ICH-PRS	0.93 [0.53-1.65]	0.81	1.24 [0.56-2.74]	0.60	0.51 [0.21-1.25]	0.14	0.57 [0.27-1.18]	0.13	0.89 [0.54-1.46]	0.64
IS-iPGS	1.22 [0.75-2.01]	0.43	1.2 [0.62-2.34]	0.59	1.34 [0.6-2.99]	0.47	1.4 [0.7-2.83]	0.34	1.35 [0.86-2.13]	0.19
IS-GRS	0.96 [0.56-1.63]	0.88	1.34 [0.66-2.74]	0.42	0.71 [0.31-1.64]	0.42	1.29 [0.67-2.49]	0.45	1.08 [0.68-1.69]	0.75
WMH-GRS	1.18 [0.74-1.9]	0.49	1.83 [0.9-3.73]	0.09	0.83 [0.4-1.71]	0.61	0.87 [0.49-1.56]	0.64	1.18 [0.79-1.76]	0.43

Supplementary Table 2b. Associations of PRS with primary and secondary endpoints adjusted for HAS-BLED

	Any Stroke		Ischemic Stroke		ICH		Major Bleed		MACE	
	HR 95% CI	**P**	**HR 95% CI**	**P**	**HR 95% CI**	**P**	**HR 95% CI**	**P**	**HR 95% CI**	**P**
AF-PRS	0.94 [0.57-1.54]	0.80	1.32 [0.68-2.55]	0.41	0.44 [0.17-1.13]	0.09	0.49 [0.23-1.03]	0.06	0.81 [0.52-1.27]	0.35
ICH-PRS	0.92 [0.52-1.65]	0.78	1.27 [0.58-2.79]	0.55	0.55 [0.21-1.43]	0.22	0.6 [0.28-1.29]	0.19	0.91 [0.55-1.5]	0.70
IS-iPGS	1.38 [0.8-2.37]	0.24	1.31 [0.63-2.72]	0.47	1.49 [0.65-3.39]	0.35	1.36 [0.71-2.6]	0.36	1.43 [0.9-2.25]	0.13
IS-GRS	1.06 [0.64-1.77]	0.82	1.42 [0.72-2.81]	0.31	0.8 [0.36-1.79]	0.58	1.29 [0.67-2.48]	0.44	1.13 [0.73-1.77]	0.58
WMH-GRS	1.22 [0.75-1.98]	0.42	1.91 [0.93-3.94]	0.08	0.82 [0.39-1.72]	0.59	0.8 [0.42-1.49]	0.48	1.18 [0.78-1.77]	0.44

Supplementary Table 3a. Interaction models for Any Stroke adjusted for CHA2DS2-VASc

PRS	arm	HR	CI_LOW	CI_HIGH	P	p_interact
AF-PRS	CONTROL	1.395	0.776	2.507	0.266	0.008
	DOAC	0.227	0.068	0.757	0.016	0.008
ICH-PRS	CONTROL	1.425	0.671	3.028	0.357	0.031
	DOAC	0.332	0.108	1.017	0.054	0.031
IS-iPRS	CONTROL	1.282	0.666	2.467	0.457	0.826
	DOAC	1.146	0.531	2.472	0.728	0.826
IS-GRS	CONTROL	1.261	0.671	2.368	0.472	0.157
	DOAC	0.612	0.274	1.364	0.23	0.157
WMH-GRS	CONTROL	1.275	0.696	2.334	0.431	0.693
	DOAC	1.051	0.494	2.235	0.898	0.693

Supplementary Table 3b. Interaction models for Any Stroke adjusted for HAS-BLED

PRS	arm	HR	CI_LOW	CI_HIGH	P	p_interact
AF-PRS	CONTROL	1.305	0.758	2.247	0.336	0.018
	DOAC	0.281	0.087	0.902	0.033	0.018
ICH-PRS	CONTROL	1.391	0.648	2.988	0.397	0.062
	DOAC	0.461	0.176	1.203	0.114	0.062
IS-iPRS	CONTROL	1.498	0.723	3.1	0.277	0.733
	DOAC	1.245	0.556	2.788	0.594	0.733
IS-GRS	CONTROL	1.425	0.757	2.682	0.273	0.141
	DOAC	0.659	0.294	1.474	0.31	0.141
WMH-GRS	CONTROL	1.364	0.741	2.513	0.319	0.549
	DOAC	1.017	0.473	2.187	0.966	0.549

Supplementary Table 4a. Interaction models for Net Clinical Benefit adjusted for CHA2DS2-VASc

PRS	**arm**	**HR**	**CI_LOW**	**CI_HIGH**	**P**	**p_interact**
AF-PRS	CONTROL	1.18	0.716	1.945	0.516	0.002
	DOAC	0.21	0.076	0.582	0.003	0.002
ICH-PRS	CONTROL	1.524	0.768	3.027	0.228	0.007
	DOAC	0.334	0.134	0.833	0.019	0.007
IS-iPRS	CONTROL	1.979	1.008	3.883	0.047	0.102
	DOAC	0.911	0.461	1.801	0.789	0.102
IS-GRS	CONTROL	1.088	0.609	1.943	0.776	0.949
	DOAC	1.056	0.526	2.122	0.878	0.949
WMH-GRS	CONTROL	1.353	0.796	2.3	0.264	0.416
	DOAC	0.964	0.516	1.801	0.909	0.416

Supplementary Table 4b. Interaction models for Net Clinical Benefit adjusted for HAS-BLED

PRS	arm	HR	CI_LOW	CI_HIGH	P	p_interact
AF-PRS	CONTROL	1.161	0.716	1.883	0.546	0.005
	DOAC	0.245	0.092	0.656	0.005	0.005
ICH-PRS	CONTROL	1.486	0.757	2.916	0.25	0.016
	DOAC	0.434	0.19	0.991	0.048	0.016
IS-iPRS	CONTROL	2.132	1.099	4.137	0.025	0.08
	DOAC	0.93	0.467	1.853	0.836	0.08
IS-GRS	CONTROL	1.183	0.662	2.116	0.571	0.823
	DOAC	1.067	0.536	2.126	0.853	0.823
WMH-GRS	CONTROL	1.359	0.795	2.324	0.261	0.398
	DOAC	0.953	0.504	1.802	0.883	0.398

Supplementary Table 5. Associations of PRS with primary and secondary endpoints in Fine-Gray Regression models

	Any Stroke		Ischemic Stroke		ICH		Major Bleed		MACE	
	HR 95% CI	P	HR 95% CI	P	HR 95% CI	P	HR 95% CI	P	HR 95% CI	P
AF-PRS	0.89 [0.48-1.64]	0.71	1.24 [0.55-2.8]	0.61	0.45 [0.19-1.05]	0.07	0.44 [0.22-0.86]	0.02	0.77 [0.45-1.31]	0.33
ICH-PRS	0.91 [0.53-1.58]	0.74	1.27 [0.71-2.26]	0.42	0.54 [0.18-1.63]	0.28	0.53 [0.23-1.23]	0.14	0.89 [0.54-1.45]	0.63
IS-iPGS	1.34 [0.75-2.38]	0.32	1.28 [0.51-3.18]	0.60	1.52 [0.7-3.3]	0.29	1.34 [0.78-2.32]	0.29	1.4 [0.9-2.2]	0.14
IS-GRS	1.06 [0.68-1.65]	0.79	1.43 [0.76-2.68]	0.27	0.78 [0.44-1.4]	0.40	1.15 [0.62-2.13]	0.67	1.1 [0.69-1.74]	0.69
WMH-GRS	1.2 [0.75-1.91]	0.45	1.76 [0.92-3.37]	0.09	0.81 [0.39-1.69]	0.58	0.88 [0.53-1.49]	0.64	1.21 [0.85-1.73]	0.30

Supplementary Table 6a. Fine-Gray Interaction models for Any Stroke

PRS	arm	HR	CI_LOW	CI_HIGH	P	p_interact
AF-PRS	CONTROL	1.23	0.69	2.21	0.48	0.04
	DOAC	0.30	0.08	1.07	0.063	0.04
ICH-PRS	CONTROL	1.23	0.70	2.16	0.47	0.19
	DOAC	0.52	0.15	1.79	0.3	0.19
IS-iPRS	CONTROL	1.36	0.60	3.11	0.46	0.95
	DOAC	1.31	0.58	2.98	0.52	0.95
IS-GRS	CONTROL	1.36	0.81	2.28	0.24	0.07
	DOAC	0.71	0.40	1.28	0.26	0.07
WMH-GRS	CONTROL	1.36	0.76	2.45	0.3	0.44
	DOAC	0.96	0.48	1.94	0.92	0.44

Supplementary Table 6b. Fine-Gray Interaction models for Net Clinical Benefit

PRS	arm	HR	CI_LOW	CI_HIGH	P	p_interact
AF-PRS	CONTROL	1.11	0.67	1.83	0.7	0.01
	DOAC	0.25	0.08	0.73	0.011	0.01
ICH-PRS	CONTROL	1.33	0.78	2.25	0.29	0.06
	DOAC	0.47	0.17	1.31	0.15	0.06
IS-iPRS	CONTROL	1.97	0.94	4.12	0.07	0.17
	DOAC	0.97	0.50	1.92	0.94	0.17
IS-GRS	CONTROL	1.12	0.62	2.03	0.7	0.91
	DOAC	1.07	0.52	2.18	0.86	0.91
WMH-GRS	CONTROL	1.42	0.90	2.26	0.13	0.24
	DOAC	0.94	0.55	1.62	0.83	0.24

Supplementary Table 7. Associations of PRS with primary and secondary endpoints adjusted by inverse probability weights to adjust for index event bias

	Any Stroke		**Ischemic Stroke**		**ICH**		**Major Bleed**		**MACE**	
	HR 95% CI	**P**	**HR 95% CI**	**P**	**HR 95% CI**	**P**	**HR 95% CI**	**P**	**HR 95% CI**	**P**
AF-PRS	1.11 [0.67-1.85]	0.68	1.52 [0.63-3.69]	0.350	0.4 [0.17-0.97]	0.04	0.44 [0.21-0.93]	0.03	1.06 [0.67-1.69]	0.793
ICH-PRS	0.69 [0.38-1.24]	0.21	1.14 [0.64-2.03]	0.665	0.37 [0.13-1.07]	0.07	0.43 [0.18-1.06]	0.07	0.76 [0.45-1.31]	0.323
IS-iPGS	2.21 [1.23-3.95]	0.01	3.55 [1.4-9]	0.008	1.16 [0.46-2.94]	0.75	1.25 [0.59-2.65]	0.56	2.24 [1.34-3.76]	0.002
IS-GRS	0.78 [0.36-1.68]	0.52	1.93 [0.91-4.09]	0.086	0.69 [0.31-1.52]	0.35	0.92 [0.43-1.97]	0.82	0.79 [0.38-1.61]	0.512
WMH-GRS	1.17 [0.65-2.11]	0.60	2.06 [0.7-6.1]	0.191	0.81 [0.41-1.57]	0.53	0.84 [0.51-1.39]	0.50	1.22 [0.74-2.04]	0.436

Supplementary Table 8a. Difference in AUC for Any Stroke at 3-years between PRS and CHA2DS2-VASc

model	**Reference**	**delta.AUC**	**p**
AF-PRS	CHA2DS2-VASc	0.0401	0.259
ICH-PRS	CHA2DS2-VASc	0.0295	0.486
IS-iPRS	CHA2DS2-VASc	0.0140	0.421
WMH-GRS	CHA2DS2-VASc	-0.0005	0.980
IS-GRS	CHA2DS2-VASc	-0.0034	0.858

Supplementary Table 8b. Difference in AUC at 3-years for Ischemic Stroke between PRS and CHA2DS2-VASc

model	reference	delta.AUC	p
WMH-GRS	CHA2DS2-VASc	0.023	0.47
ICH-PRS	CHA2DS2-VASc	-0.005	0.73
AF-PRS	CHA2DS2-VASc	-0.010	0.55
IS-iPRS	CHA2DS2-VASc	-0.012	0.28
IS-GRS	CHA2DS2-VASc	-0.021	0.23

Supplementary Table 8c. Difference in AUC at 3-years for ICH between PRS and HAS-BLED

model	reference	delta.AUC	p
AF-PRS	HASBLED	0.092	0.211
ICH-PRS	HASBLED	0.045	0.337
IS-GRS	HASBLED	0.019	0.601
IS-iPRS	HASBLED	0.010	0.792
WMH-GRS	HASBLED	-0.003	0.909

Supplementary Table 8d. Difference in AUC at 3-years for Major Bleeds between PRS and HAS-BLED

model	reference	delta.AUC	p
AF-PRS	HASBLED	0.094	0.111
ICH-PRS	HASBLED	0.059	0.120
IS-iPRS	HASBLED	0.017	0.540
WMH-GRS	HASBLED	0.004	0.850
IS-GRS	HASBLED	-0.003	0.848

Supplementary Table 8e. Difference in AUC at 3-years for net clinical benefit between PRS and CHA2DS2-VASc

model	reference	delta.AUC	p
AF-PRS	CHA2DS2	0.101	0.029
ICH-PRS	CHA2DS2	0.060	0.211
IS-iPRS	CHA2DS2	0.034	0.443
WMH-GRS	CHA2DS2	-0.003	0.884
IS-GRS	CHA2DS2	-0.009	0.165

Supplementary Table 9a-d. DOAC absolute risk reduction by AF-PRS risk tertile

Any Stroke				
PRS Tertile + Treatment Group	N	Cases	Event Rate	DOAC ARR
CONTROL + Low PRS	24	4	16.7	-6.1% [-31.2,18.9]
DOAC + Low PRS	19	2	10.5	
CONTROL + High PRS	20	4	20.0	-20% [-42.3,2.3]
DOAC + High PRS	22	0	0.0	
Ischemic Stroke				
PRS Tertile + Treatment Group	N	Cases	Event Rate	DOAC ARR
CONTROL + Low PRS	24	3	12.5	-12.5% [-30.4,5.4]
DOAC + Low PRS	19	0	0	
CONTROL + High PRS	20	4	20	-20% [-42.3,2.3]
DOAC + High PRS	22	0	0	
ICH				
PRS Tertile + Treatment Group	N	Cases	Event Rate	DOAC ARR
CONTROL + Low PRS	24	1	4.2	+6.4% [-34.3, 27.0]
DOAC + Low PRS	19	2	10.5	
CONTROL + High PRS	20	0	0.0	0% [-4.8,4.8]
DOAC + High PRS	22	0	0.0	
Major Adverse Cardiovascular Events				
PRS Tertile + Treatment Group	N	Cases	Event Rate	DOAC ARR
CONTROL + Low PRS	24	5	20.8	+0.2% [-29,29.4]
DOAC + Low PRS	19	4	21.1	
CONTROL + High PRS	20	5	25.0	-25% [-48.7,-1.3]
DOAC + High PRS	22	0	0.0	

Supplementary Table 10a-d. DOAC absolute risk reduction by IS-iPRS risk tertile

Any Stroke

PRS Tertile + Treatment Group	N	Cases	Event Rate	DOAC ARR
CONTROL + Low PRS	19	2	10.5	-2.2% [-24.6,20.2]
DOAC + Low PRS	24	2	8.3	
CONTROL + High PRS	22	5	22.7	-7.7% [-36,20.5]
DOAC + High PRS	20	3	15.0	

Ischemic Stroke

PRS Tertile + Treatment Group	N	Cases	Event Rate	DOAC ARR
CONTROL + Low PRS	19	2	10.5	-10.5% [-29,8.0]
DOAC + Low PRS	24	0	0	
CONTROL + High PRS	22	5	22.7	-22.7% [-45.0,-0.4]
DOAC + High PRS	20	0	0	

ICH

PRS Tertile + Treatment Group	N	Cases	Event Rate	DOAC ARR
CONTROL + Low PRS	19	0	0.0	+8.3% [-7.4,24.1]
DOAC + Low PRS	24	2	8.3	
CONTROL + High PRS	22	0	0.0	+15% [-5.4,35.4]
DOAC + High PRS	20	3	15.0	

Major Adverse Cardiovascular Events

PRS Tertile + Treatment Group	N	Cases	Event Rate	DOAC ARR
CONTROL + Low PRS	19	2	10.5	+2% [-21.9,25.8]
DOAC + Low PRS	24	3	12.5	
CONTROL + High PRS	22	8	36.4	-21.4% [-51.6,8.9]
DOAC + High PRS	20	3	15.0	

Appendix 8

Pathway-specific polygenic risk scores reveal underlying mechanisms of cerebral small vessel disease

Supplement

Table of contents

Supplementary Methods **2**

MRI Sample exclusions 2

MRI Acquisitions and WMH calculation 3

Genotyping collection and quality control 4

PRS Methods 5

PRSet Enrichment 5

Clinical Outcomes Secondary Analysis 5

Hypertension definition in UKBB 8

Acquisition and transformation of NDI metrics in i-Share 8

References 10

Supplementary Figures **13**

Supplementary Tables **15**

Supplementary Methods

MRI Sample exclusions

UK Biobank MRI

We also excluded individuals who had an MRI prior to the reception date of the WMH GWAS from which we derived the risk variants and weights (January 26th, 2020) to avoid sample overlap bias (1). We determined the UKBB-MRI sample on June 15th, 2025. As of this date, 16,320 individuals of European ancestry had available white matter hyperintensity volume recorded after January 26th, 2020. For related pairs (2nd degree relatives, kinship > 0.0844), we randomly selected one individual to remove. We removed individuals with prevalent neurological conditions prior to MRI based on ICD-10 codes provided by field codes 41270 and 41280: Alzheimer's Disease (ICD 10 codes G30-32), Parkinson's disease (ICD-10 codes, G20), Stroke (I60-64), and multiple sclerosis (G35-37). We also additionally added prevalent cases for those with algorithmically-defined outcomes for stroke (field 42006), Parkinson's disease (field 42032), and dementia (field 42018).

From the sample 16,360 we had the following exclusions:

- Missing HRC imputed genotyped data, n = 35
- In a related pair, n = 155
- Prevalent neurological conditions, n = 276
- Missing eTIV, n = 574

Final sample: 15,320 participants

3C-Dijon

1,924 in 3C-Dijon underwent MRI

Exclusions:

- Low-quality MRI, n = 10
- No WMH and white matter volume mask, n = 113
- No viable genetic data, n = 168
- Prevalent stroke or dementia or no data on stroke or dementia history, n = 106
- Brain tumor, n = 6
- 3rd degree relatedness, n = 78

Final Sample: 1,443 participants

i-Share

For the present study, the population comprised i-Share participants taking part in both MRi-Share and bio-Share and for whom brain MRI and genome-wide genotype data passed quality control filters. Of 1,954 participants who took part in MRi-Share, 1,856 had a usable (after quality control, QC) brain MRI including measures of subcortical brain structures. Genome-wide genotype data were available for 1,862 individuals. In total, 1,784 participants had both high quality brain MRI and genome-wide genotype data available.
Additionally:

- Seven individuals were excluded with a diagnosis or suspicion of multiple sclerosis or radiologically isolated syndrome
- 151 were excluded due to non-European ancestry

- 48 were excluded due to 3rd degree relatedness

The final i-Share sample had 1,578 eligible participants (mean age=22.1±2.3 years, 71.7% female).

Memento

The Memento study is a memory clinic-based cohort aimed to investigate the natural history of AD and related disorders in a large group of participants with MCI or subjective cognitive complaints. Memento participants were recruited from 26 French memory clinics in 2011-2014, with a follow-up every 6-12 months over 5 years (2). We excluded those with missing WMH volume or genetic data, with prevalent stroke, dementia or Parkinson's disease, or a 3rd degree relative in the sample, resulting in 1,831 eligible participants (mean-age=71.1±8.5 years, 62.2% female).

2,246 individuals underwent MRI

Exclusions:

- Missing WMH volume, n = 142
- Extreme intracranial volume, n = 1
- No viable genetic data, n = 170
- Prevalent dementia, n = 18
- Prevalent stroke, n = 79
- Parkinson's disease, n = 4
- Relatedness, n = 1

Final sample : 1,831 participants

MRI Acquisitions and WMH calculation

UK Biobank

Details of the UK Biobank MRI protocol and documentation is publically available (https://www.fmrib.ox.ac.uk/ukbiobank/) and image-derived phenotyps are fully described in a previous publication (3). In brief, the data were acquired with a standard Siemens Skyra 3T 32-channel head coil. Total volume of WMH was estimated via T_1-weighted MPRAGE and T_2-weighted FLAIR volumes. The total volume of WMH calculated using T_1-weighted MPRAGE and T_2-weighted FLAIR volumes segmented by the BIANCA tool (4).

3C-Dijon

MRI scans in 3C-Dijon were acquired from a 1.5-Tesla Magnetom scanner (Siemens, Erlangen, Germany) (5). T1- and T2-weighted images of each subject were aligned to each other using the AIR package. These images were analyzed with the optimized Voxel-Based Morphometry (VBM) protocol, using Statistical Parametric Mapping 99 (SPM99) that were modified in order to take into account the structural characteristics of the aged brain. Fully automated image processing software was developed to detect, measure, and localize white matter hyperintensities (WMH) (6).

i-Share

1,832 i-Share participants participated in the MRi-Share sub-study and underwent MRI acquisition protocol designed to closely emulate that which was done in UK Biobank (7). T1-weighted and T2-weighted FLAIR structural scans were acquired with a Siemens 3T PRisma scanner with a 64-channels head coil between November 2015 and November 2017. Due to low WMH volume in young adults, WMH volume was derived with the SHIVA-WMH tool developed to improve detection of small and sparse WMH (8).

Memento

In Memento, 86% of participants had a 3 Telsa or 1.5 T MRI scan at their respective memory clinics (2). Images were harmonized by the Center for Automated Treatment of Images (CATI) (9), and standard acquisitions were applied according to a systematic qualification procedure ensuring parameter uniformity and image quality. MRI sequences for T1-weighted and T2-weighted FLAIR were conducted to ensure compatibility with Alzheimer's Disease Neuroimaging Initiative (ADNI) protocol. WMH volume was measured with White Matter Hyperintensities Automated Segmentation Algorithm software (10) and quantitatively verified by two trained physicians.

Genotyping collection and quality control

UK Biobank

Quality control and imputation with the Haplotype Reference Consortium reference panel are fully described elsewhere (11). We were granted access to UK Biobank data through application number 94113. We used previously generated relatedness variables available of the UK Biobank Research Analysis Platform to identify related pairs (kinship ≥ 0.0844) and randomly selected individuals in each pair for exclusion (n=155).

3C

For 3C participants, genome-wide genotyping was performed at the Centre National de Génotypage in Evry (France) using the Illumina Human610Iquad BeadChips. Standard quality control procedures were initially conducted on raw genotyped data. This included removing SNPs due to missingness (<98% call rate), low minor allele frequency (MAF) (<1%), non-autosomal location, and deviation from Hardy-Weinberg equilibrium ($p<0.001$). Participants were removed with discordant sex information, excessive missingness (<95% call rate), high heterozygosity (±3SD), or divergent ancestry (±6SD) (12). Additionally, sample pairs were identified with 3rd-degree relatedness with KING software (kinship > 0.0625), and were removed. The raw genotyped data was then imputed via the Michigan server using the Haplotype Reference Consortium (HRC) (13).

i-Share

Genome-wide genotyping of 1 872 i-Share participants was performed using the Affymetrix Precision Medicine Axiom Array at McGill Genome Center (Canada). After quality control, genotype data were available for 1 862 participants (7 participants were removed due to sex discrepancies, 2 participants who appeared to be duplicates but not twins, and one participant with a kinship coefficient >0.0625 (third degree related) with more than 20 other participants, suggesting a possible sample contamination (KING software)).[19] After applying standard quality control procedures (SNP call rate <98%, Hardy-Weinberg Equilibrium $p<0.001$), we imputed the genotypes on the Haplotype Reference Consortium (HRC) reference panel.[3].

Memento

Pre-imputation QC included removal of SNPs with MAF<0.01, callrate<0.98 and HWE<0.001, removal of samples with callrate <0.05, heterozygosity beyond 3SD, failed sex-check using genotype data of X-chromosome, related sample based on IBD (pi_hat>0.1875), PCA outliers beyond 6SD of PC1 and PC2. The raw genotyped data was then imputed via the Michigan server using the Haplotype Reference Consortium (HRC) (13). Additional related pairs were identified with 3rd-degree relatedness with KING software (kinship > 0.0625). We randomly removed one individual from each related pair.

PRS Methods

In clumping and thresholding, SNPs are first clumped to account for LD, selecting the most significant within ±250kb, and removing any variants correlated with the index SNP (r^2>0.1). A range of p-value thresholds (p_t=5x10^{-8}, 1x10^{-6}, 1x10^{-4}, 0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 1) is then applied to shrink effect sizes for SNPs with a p-value over the threshold to zero. Alternatively, PRS-CS is a Bayesian approach that utilizes a continuous shrinkage prior on SNP effect sizes to directly model LD and limit loss of information (14). PRS-CS requires tuning of the global shrinkage parameter (ϕ) representing prior knowledge about the sparseness of the underlying genetic architecture. The starting ϕ value can be manually set or tuned 'automatically', thus, we used φ=1, 1x10^{-2}, 1x10^{-4}, 1x10^{-6}, 1x10^{-8}, or 'auto'.

PRSet Enrichment

PRSet relies on empirical, "competitive" P-values derived from a permutation-based procedure to determine pathway enrichment. PRSet is applied on clumped SNPs (r^2<0.1) without any p-value thresholding. This is to ensure that all genes in the pathway are represented regardless of their underlying association to the trait. The permutation-based procedure also accounts for the number of SNPs included in a given pathway, so as not to privilege pathway enrichment towards pathways with a larger number of genes.

In brief, the method is summarized in the PRSet paper:

1. A "background" pathway containing all genic SNPs is constructed, and clumping is performed within this pathway. For pathways with m SNPs, N null pathways are generated by randomly selecting m "independent" SNPs from the "background" pathway.
2. The competitive P-value can then be calculated as

$$competitive\ P-value = \frac{\sum_{n=1}^{N} I(P_n < P_o) + 1}{N + 1}$$

Where: I(.) is an indicator function, taking a value of 1 if the association P-value of the observed gene set (P0) is larger than the one obtained from the nth null set (Pn), and 0 otherwise. A pseudo-count of 1 is added to the numerator and denominator to avoid competitive P-values of 0 and conservatively counting the observed gene set as 1 potential null set [54]. One consideration of this permutation procedure is that the smallest achievable competitive P-value is 1/(N+1), which can lead to difficulties in ranking highly significant gene sets.

This method is recommended by the authors over standard regression procedures and multiple testing correction.

Clinical Outcomes Secondary Analysis

Below you will find study specific exclusions for the clinical outcome secondary analysis, and outcome definitions. Each study and outcome also required specific modeling in order to fit cox proportional hazard models that satisfied the cox proportional hazard assumption.

UKBB

For analysis of clinical outcomes in UK Biobank, we excluded participants with brain MRI to avoid overlap with the WMH GWAS and also with the UKBB-MRI sample used to identify enriched WMH ps-PRS. We also exclude those of non-European ancestry, and those in a related pair (2nd degree relative). For analyses of stroke or dementia outcomes, we exclude individuals with prevalent conditions. We also remove individuals with self-reported stroke. Individuals were censored at clinical

outcome, death, or end-of-follow-up was considered as of May 31, 2022 – the latest date where hospital records are mostly complete from Wales, Scotland, and England. We ascertained stroke and dementia outcomes using algorithmically-defined stroke and dementia outcomes to identify incident stroke cases defined during follow-up with UK Biobank fields 42006-42010 (any stroke, IS, and ICH), and 42018-42022 (dementia, AD, and vascular dementia) (https://biobank.ndph.ox.ac.uk/showcase/ukb/docs/alg_outcome_main.pdf).

Exclusions:

- European ancestry, n = 93,579
- Relatedness, n = 3,535
- Brain MRI (as of June 15th, 2025), n = 49,810

Stroke-specific exclusions:

- Prevalent stroke, n = 6,046
- Self-reported stroke, n = 28

Sample size for incident stroke analyses: N=349,283

- Undefined stroke cases removed from subtype-specific analysis: n = 587

Dementia-specific exclusions:

- Prevalent dementia, n = 176
- Self-reported incident dementia, n = 1

Sample size for incident dementia analyses: N= 355,180

- Undefined dementia cases removed from subtype-specific analysis: n = 4,009

End-of-follow-up was considered as of May 31, 2022 – the latest date where hospital records are mostly complete from Wales, Scotland, and England.

Stroke Modeling in UKBB

- **Any Stroke:** cox proportional hazard models were stratified by sex and geno-chip in order to satisfy proportional hazard assumptions, allowing baseline risks to vary across these strata. Models were further adjusted by 10 genetic principal components.
- **ICH or IS:** cox proportional hazard models were stratified by sex. Models were adjusted for genotype chip and 10 principal components.

Dementia Modeling in UKBB

- **Any Dementia:** cox proportional hazard models were stratified by sex, and adjusted for genotype chip, 10 PCs, and number of APOE-ε2 and APOE-ε4 alleles (0, 1 or 2)
- **AD:** cox proportional hazard models were adjusted for sex, genotype chip, 10 PCs, and number of APOE-ε2 and APOE-ε4 alleles (0, 1 or 2)
- **VMD:** cox proportional hazard models were stratified by sex, and adjusted for genotype chip, 10 PCs, and number of APOE-ε2 and APOE-ε4 alleles (0, 1 or 2)

Three Cities (3C)

In 3C, an expert panel reviewed potential stroke cases identified at each follow-up visit with participants or informants for deceased participants (15). Expert panels of neurologists reviewed all available clinical information, confirming diagnosis of stroke and its subtype (ischemic or haemorrhagic) according to the International Classification of Diseases – 10th Edition (16). Diagnosis of dementia followed a three-step procedure at baseline and at each follow-up visit (17). First, all participants underwent a battery of

neuropsychological tests at baseline and each follow-up visit by a trained psychologist. Second, a neurologist examined all the participants in Montpellier and in Bordeaux at baseline. In Dijon, participants screening positive for dementia based on their neuropsychological performance underwent further clinical examination. At each follow-up, all participants with suspected dementia were examined by a neurologist to establish provisional diagnosis. We combined vascular and mixed dementia diagnoses to capture cases with a more substantial vascular component.

Individuals were censored at death, clinical outcome, or end of follow-up. If an individual died 2.5 years after their last visit, they were censored at the date of their last visit since it is unknown if they had a stroke or developed dementia during this period.

Exclusions:

Full sample: 9,294

Stroke

- No viable imputed genotypes, n = 2,805
- Missing stroke information, n = 294 (could be due to just insufficient follow-up)
- Relatedness, 3rd degree, n = 177
- Prevalent stroke, n = 273

N = 5,745

Dementia

- No viable imputed genotypes, n = 2,805
- Relatedness, 3rd degree, n = 181
- Prevalent dementia, n = 121
- Insufficient follow-up, n = 304

N = 5,883

Stroke Modeling in 3C

- **Any Stroke:** cox proportional hazard models adjusted for sex, study centre, and 4 genetic principal components
- **ICH:** cox proportional hazard models were stratified by sex, and adjusted for study centre, and 4 PCs
- **IS:** cox proportional hazard models adjusted for sex, study centre, and 4 genetic principal components

Dementia Modeling in 3C

- Cox proportional hazard models for dementia outcomes were all stratified by study centre, and adjusted for sex, 4 PCs, and dosage of APOE-ε2 and APOE-ε4

Memento

In Memento, trained neurologists administered a battery of neuropsychological tests at baseline, and assessed dementia status at each follow-up visit (2). In both cohorts, an independent panel of expert neurologists reviewed and validated all possible dementia cases according to the *Diagnostic and Statistical Manual of Mental Disorders* (Fourth Revision) criteria. We combined vascular and mixed dementia diagnoses to capture cases with a more substantial vascular component.

Individuals were censored at death, dementia, or end of follow-up. If an individual died one year after their last visit, they were censored at the date of their last visit since it is unknown if they developed dementia during this period.

Exclusions

Full sample: 2,323

- Missing phenotype information, n = 119
- No viable imputed genotypes, n = 87
- Prevalent dementia, n = 19
- Insufficient follow-up, n = 34

N = 2,064

Study-specific Modeling in Memento

In Memento we compared random effect vs. fixed effects for site ID since there were 33 of them, in order to control the number of parameters in the model. We compared both modelling approaches using AIC, for which mixed effects consistently had better fit. For Memento we used mixed effect cox proportional hazard models (18) using the coxme R package (19) allowing for a random effect based on memory-clinic site, of which were there were 33. Models were further adjusted for sex, 7 PCs, and dosage of APOE-ε2 and APOE-ε4.

Hypertension definition in UKBB

Where available, we used automated DBP and SBP readings at the imaging assessment (field codes 4079 and 4080, instance 2) to ascertain blood pressure at baseline. We also used self-reported medical conditions data at the imaging assessment to ascertain whether an individual was taking blood pressure medication – field codes 6153 and 6177 for females and males respectively. We classified individuals as hypertensive if they had an automated SBP $\geq$ 140 mmHg or DBP $\geq$ 90 mmHg or if they self-reported blood pressure medication. Individuals were excluded if they did not have blood pressure readings and if they also did not report taking blood pressure medication or did not self-report any medications. A total of 2,800 individuals were excluded from this analysis.

In total, 4,306 self-reported taking blood pressure medication at the time of the imaging assessment, 7,469 had SBP $\geq$ 140 mmHg or DBP $\geq$ 90 mmHg.

Acquisition and transformation of NDI metrics in i-Share

Acquisition, derivation, and transformation of the global NDI and six brain ROIs are fully described in previous work (20). In the original work, one global measure of NDI was generated across the full white matter, in addition to 27 regional markers. We select brain ROIs associated with genetic-risk variants for WMH (combined in a GRS), and also known to have highest frequency of WMH later in life. To normalize distributions of NDI metrics, we applied a rank-based inverse normal transformation.

T1-weighted images were processed using FreeSurfer version 6.0 (http://surfer.nmr.mgh.harvard.edu/) to derive volumetric segmentation and quantitative volumetric measures (21,22). The estimated total intracranial volume (eTIV) generated by FreeSurfer was used as a proxy for head size.

Preprocessing of the dMRI data included correction of susceptibility-induced and eddy current-induced distortions as well as head motion using the FSL Eddy tool (www.fmrib.ox.ac.uk/fsl). Neurite density index (NDI) from the NODDI model (23) was derived using voxel-wise model fitting. This was performed with the AMICO (accelerated microstructure via convex optimization) tool (24).

Subsequently, the standard tract-based spatial statistics (TBSS) framework in FSL was applied to project all estimated dMRI parameters onto a template FA skeleton (25). Image-derived phenotypes were generated using the WM skeleton and atlas-based regions of interest (ROIs). Mean values were calculated across voxels within the WM skeleton, and ROI-specific means were obtained from labels corresponding to well-known white matter tracts provided by the JHU ICBM DTI 81 atlas (26) available in FSL. Six bilateral ROIs from this atlas were selected to generate regional phenotypes. For the remaining JHU atlas ROIs, the left and right hemisphere labels were merged prior to computing mean voxel values.

References

1. Choi SW, Mak TSH, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. Nat Protoc. 2020 Sep;15(9):2759–72.

2. on behalf of the MEMENTO cohort Study Group, Dufouil C, Dubois B, Vellas B, Pasquier F, Blanc F, et al. Cognitive and imaging markers in non-demented subjects attending a memory clinic: study design and baseline findings of the MEMENTO cohort. Alzheimers Res Ther. 2017 Dec;9(1):67.

3. Alfaro-Almagro F, Jenkinson M, Bangerter NK, Andersson JLR, Griffanti L, Douaud G, et al. Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. NeuroImage. 2018 Feb 1;166:400–24.

4. Griffanti L, Zamboni G, Khan A, Li L, Bonifacio G, Sundaresan V, et al. BIANCA (Brain Intensity AbNormality Classification Algorithm): A new tool for automated segmentation of white matter hyperintensities. NeuroImage. 2016 Nov 1;141:191–205.

5. Godin O, Dufouil C, Maillard P, Delcroix N, Mazoyer B, Crivello F, et al. White Matter Lesions as a Predictor of Depression in the Elderly: The 3C-Dijon Study. Biol Psychiatry. 2008 Apr 1;63(7):663–9.

6. Maillard P, Delcroix N, Crivello F, Dufouil C, Gicquel S, Joliot M, et al. An automated procedure for the assessment of white matter hyperintensities by multispectral (T1, T2, PD) MRI and an evaluation of its between-centre reproducibility based on two large community databases. Neuroradiology. 2008 Jan 1;50(1):31–42.

7. Tsuchida A, Laurent A, Crivello F, Petit L, Joliot M, Pepe A, et al. The MRi-Share database: brain imaging in a cross-sectional cohort of 1870 university students. Brain Struct Funct. 2021 Sep 1;226(7):2057–85.

8. Tsuchida A, Boutinaud P, Verrecchia V, Tzourio C, Debette S, Joliot M. Early detection of white matter hyperintensities using SHIVA-WMH detector. Hum Brain Mapp. 2024;45(1):e26548.

9. Operto G, Chupin M, Batrancourt B, Habert MO, Colliot O, Benali H, et al. CATI: A Large Distributed Infrastructure for the Neuroimaging of Cohorts. Neuroinformatics. 2016 Jul 1;14(3):253–64.

10. Samaille T, Fillon L, Cuingnet R, Jouvent E, Chabriat H, Dormont D, et al. Contrast-Based Fully Automatic Segmentation of White Matter Hyperintensities: Method and Validation. PLOS ONE. 2012 Nov 12;7(11):e48953.

11. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018 Oct;562(7726):203–9.

12. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. Nat Protoc. 2010 Sep;5(9):1564–73.

13. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. Nat Genet. 2016 Oct;48(10):1284–7.

14. Ge T, Chen CY, Ni Y, Feng YCA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. Nat Commun. 2019 Apr 16;10(1):1776.

15. Samieri C, Féart C, Proust-Lima C, Peuchant E, Tzourio C, Stapf C, et al. Olive oil consumption, plasma oleic acid, and stroke incidence. Neurology. 2011 Aug 2;77(5):418–25.

16. World Health Organization. International Statistical Classification of Diseases and Related Health Problems 10th revision [Internet]. [cited 2024 Dec 4]. Available from: https://icd.who.int/browse10/2016/en

17. Vascular Factors and Risk of Dementia: Design of the Three-City Study and Baseline Characteristics of the Study Population. Neuroepidemiology. 2003;22(6):316–25.

18. Ripatti S, Palmgren J. Estimation of Multivariate Frailty Models Using Penalized Partial Likelihood. Biometrics. 2000 Dec 1;56(4):1016–22.

19. Therneau T. coxme: Mixed Effects Cox Models. 2024.

20. Le Grand Q, Tsuchida A, Koch A, Imtiaz MA, Aziz NA, Vigneron C, et al. Diffusion imaging genomics provides novel insight into early mechanisms of cerebral small vessel disease. Mol Psychiatry. 2024 Nov;29(11):3567–79.

21. Fischl B. FreeSurfer. NeuroImage. 2012 Aug 15;62(2):774–81.

22. Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, et al. Whole Brain Segmentation. [cited 2025 Sep 30]; Available from: https://www.cell.com/neuron/abstract/S0896-6273(02)00569-X

23. Zhang H, Schneider T, Wheeler-Kingshott CA, Alexander DC. NODDI: Practical *in vivo* neurite orientation dispersion and density imaging of the human brain. NeuroImage. 2012 Jul 16;61(4):1000–16.

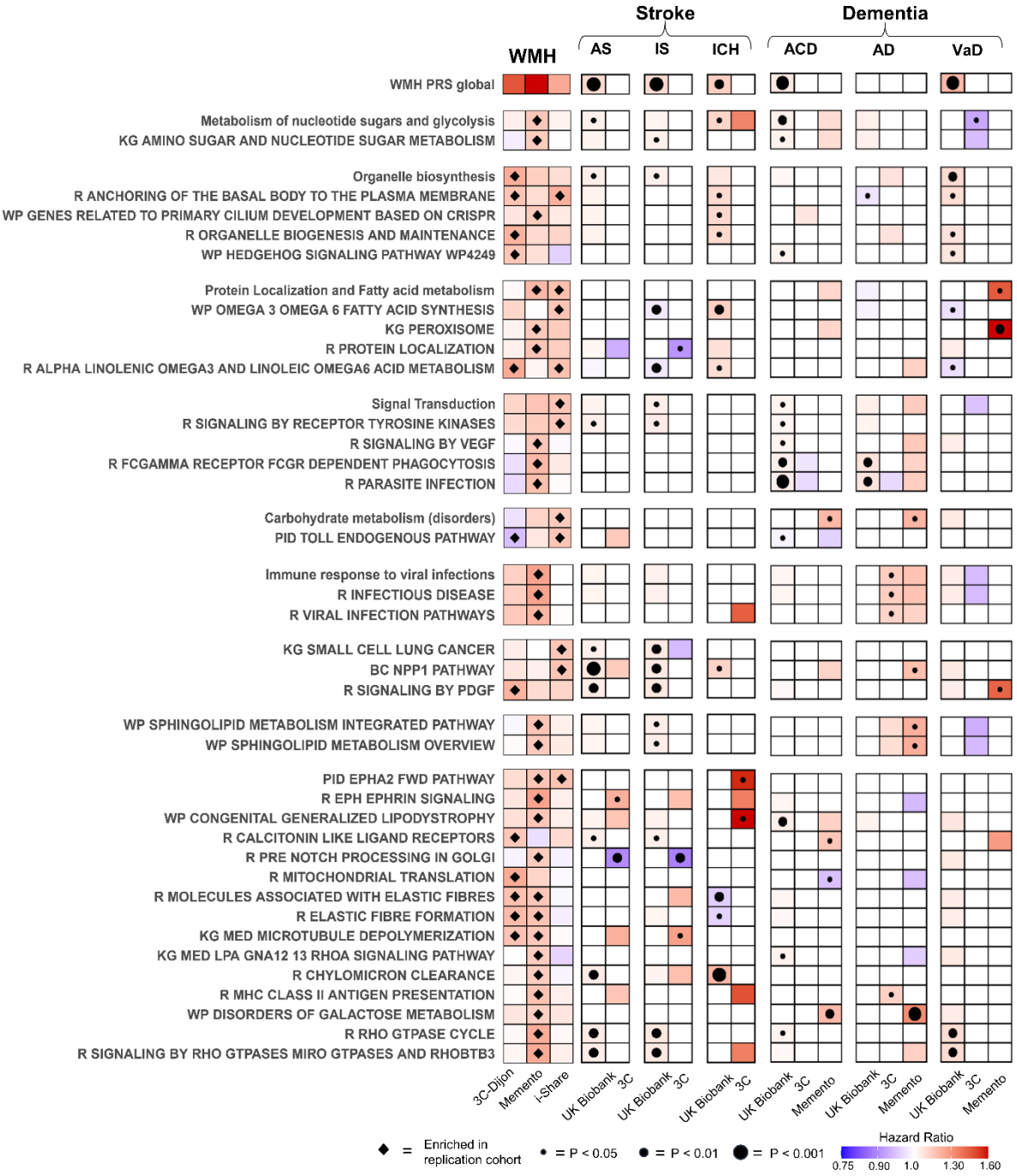
24. Daducci A, Canales-Rodríguez EJ, Zhang H, Dyrby TB, Alexander DC, Thiran JP. Accelerated Microstructure Imaging via Convex Optimization (AMICO) from diffusion MRI data. NeuroImage. 2015 Jan 15;105:32–44.

25. Smith SM, Jenkinson M, Johansen-Berg H, Rueckert D, Nichols TE, Mackay CE, et al. Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data. NeuroImage. 2006 Jul 15;31(4):1487–505.

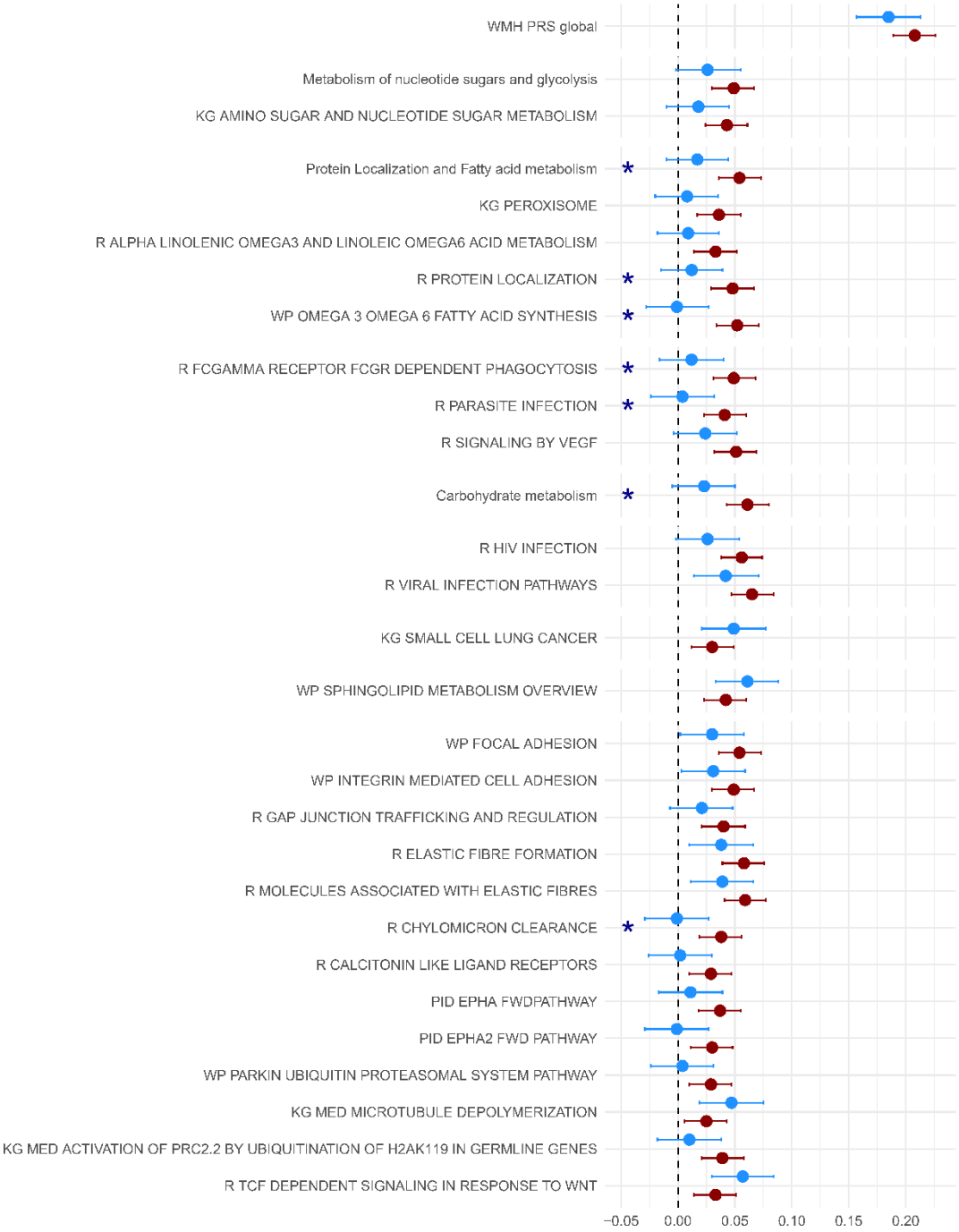
26. Mori S, Oishi K, Jiang H, Jiang L, Li X, Akhter K, et al. Stereotaxic white matter atlas based on diffusion tensor imaging in an ICBM template. NeuroImage. 2008 Apr 1;40(2):570–82.

Supplementary Figures

Supplementary Figure 1. Associations of all WMH-psPRS included in the secondary analysis in Cox proportional hazard models. All models are adjusted for adjusted for sex, genetic PCs and other study-specific covariates. Dementia was additionally adjusted by *APOE-ε2* and *ε4* dosage.



Supplementary Figure 2. Effect modification by hypertension in the association of WMH-(ps)PRS with WMH volume. Linear regression models for WMH volume were adjusted baseline covariates: age at MRI, sex, 10PCs, genotyping chip and total intracranial volume. We show WMH ps-PRS with a heterogeneity P-value < 0.3. * indicates nominal significance (p<0.05



Supplementary Tables

Supplementary Table 1. *WMH-PRS global 10-fold cross validation results in UKBB MRI sample. 16 candidate models were tested using PRS-cs and C+T methods.*

PRS-MODEL	BETA	CI-LOW	CI-HIGH	P	R2 improvement vs. base model (%)	RSME 10-fold CV
PRS-CS phi=auto	0.2036	0.19	0.217	1.8E-193	4.08	0.8316
PRS-CS phi=1e-04	0.2021	0.189	0.215	1.1E-190	4.02	0.8319
PRS-CS phi=1e-06	0.1841	0.171	0.197	1.3E-159	3.37	0.8358
PRS-CS phi=1e-02	0.1681	0.155	0.182	1.3E-129	2.74	0.8395
C+T pt=0.01	0.1525	0.139	0.166	1.2E-108	2.3	0.8421
C+T pt=0.1	0.1498	0.136	0.163	4.6E-103	2.18	0.8429
C+T pt=1e04	0.1467	0.133	0.16	2.0E-101	2.14	0.8431
C+T pt=0.2	0.1458	0.132	0.159	1.6E-97	2.06	0.8436
C+T pt=0.3	0.1445	0.131	0.158	1.4E-95	2.02	0.8438
C+T pt=0.4	0.1435	0.13	0.157	2.9E-94	1.99	0.844
C+T pt=0.5	0.1426	0.129	0.156	4.9E-93	1.97	0.8441
C+T pt=1	0.1422	0.129	0.156	1.5E-92	1.95	0.8442
PRS-CS phi=1	0.1369	0.123	0.15	8.0E-86	1.81	0.8451
C+T pt=1e06	0.1351	0.122	0.148	5.3E-86	1.81	0.845
C+T pt=5e08	0.1347	0.121	0.148	1.5E-85	1.8	0.8451
PRS-CS phi=1e-08	0.134	0.121	0.147	1.3E-84	1.79	0.8452

Supplementary Table 2. *Overview of ps-PRS clusters with gene ontology terms, number of genes, average correlation between pathways, and list of pathways within clusters*

General Pathway Name	Pathways	Number of Genes	GO-MF	GO-BP	GO-CC	Human Phenotypes
Metabolism and biosynthesis of nucleotide sugars and glycolysis	KEGG AMINO SUGAR AND NUCLEOTIDE SUGAR METABOLISM; KEGG GALACTOSE METABOLISM; KEGG MEDICUS REFERENCE PINK PARKIN MEDIATED AUTOPHAGOSOME FORMATION; WP AEROBIC GLYCOLYSIS; KEGG MEDICUS VARIANT MUTATION INACTIVATED OPTN TO PINK PARKIN MEDIATED AUTOPHAGOSOME FORMATION; KEGG MEDICUS VARIANT MUTATION INACTIVATED P62 TO PINK PARKIN MEDIATED AUTOPHAGOSOME FORMATION; WP GLYCOLYSIS AND GLUCONEOGENESIS	111	catalytic activity monosacchride binding	carbohydrate metabolic process pyruvate metabolic process	extracellular exosome	abnormal enzyme concentration or activity
Organelle Biosynthesis and Maintenance (cilium)	WP GENES RELATED TO PRIMARY CILIUM DEVELOPMENT BASED ON CRISPR; WP HEDGEHOG SIGNALING PATHWAY WP47; REACTOME ANCHORING OF THE BASAL BODY TO THE PLASMA MEMBRANE; REACTOME CILIUM ASSEMBLY; REACTOME ORGANELLE BIOGENESIS AND MAINTENANCE; WP BARDET BIEDL SYNDROME; WP HEDGEHOG SIGNALING PATHWAY WP4249; REACTOME RUNX2 REGULATES OSTEOBLAST DIFFERENTIATION	415	protein binding	cilium organization microtubule-based process organelle assembly	cilium microtuble cytoskeleton proton-trnasporting ATP synthase complex	polydactyl
Protein Localization and Import and Fatty Acid metabolism (peroxisomes)	KEGG ALPHA LINOLENIC ACID METABOLISM; WP OMEGA 3 OMEGA 6 FATTY ACID SYNTHESIS; WP EICOSANOID METABOLISM VIA CYCLOOXYGENASES COX; REACTOME PEROXISOMAL LIPID METABOLISM; REACTOME PEROXISOMAL PROTEIN IMPORT; REACTOME PROTEIN LOCALIZATION; REACTOME ALPHA LINOLENIC OMEGA3 AND LINOLEIC OMEGA6 ACID METABOLISM; KEGG PEROXISOME; WP EICOSANOID METABOLISM VIA LIPOOXYGENASES LOX	247	oxidoreductase activity protein transporter activity calcium-dependent phospholipase A2 activity	fatty acid metabolic process protein transmembrane import into intracellular organelles	microbody mitochondrion	Abnormal circulating long-chain-fatty-acid concentration
Inflammatory Signaling	REACTOME COENZYME A BIOSYNTHESIS; KEGG MEDICUS REFERENCE CA2 PLCD ITPR SIGNALING PATHWAY; REACTOME VITAMIN B5 PANTOTHENATE METABOLISM; WP RESISTIN AS A REGULATOR OF INFLAMMATION	55	phosphatidylinositol phospholipase C activity	phosphatidylinositol-mediated signaling coenzyme A metabolic process	cytoplasm	Autoimmunity
Extracellular Matrix	REACTOME EXTRACELLULAR MATRIX ORGANIZATION; WP MIRNA TARGETS IN ECM AND MEMBRANE RECEPTORS; PID AVB3 INTEGRIN PATHWAY; REACTOME LAMININ INTERACTIONS; REACTOME SIGNALING BY PDGF; KEGG SMALL CELL LUNG CANCER; REACTOME ASSEMBLY OF COLLAGEN FIBRILS AND OTHER MULTIMERIC STRUCTURES; BIOCARTA NPP1 PATHWAY	438	extracellular matrix structural constituent	extracellular matrix organization cell adhesion response to endogenous stimulus	extracellular matrix extracellular space	Atrophic scars
Signal Transduction	REACTOME SIGNALING BY VEGF; REACTOME SIGNALING BY RECEPTOR TYROSINE KINASES; REACTOME FCGAMMA RECEPTOR FCGR DEPENDENT PHAGOCYTOSIS; REACTOME PARASITE INFECTION	623	signaling receptor binding exyme bindig protein kinase activity	cell surface receptor protein tyrosine kinase signaling pathway response to endogenous stimulus protein phosphorylation immune system process	cell periphery protein-transporting V-type ATPase complex	autosomal dominant inheritance
Non-homologous end-joining	KEGG NON HOMOLOGOUS END JOINING; REACTOME EARLY PHASE OF HIV LIFE CYCLE; KEGG MEDICUS REFERENCE NON HOMOLOGOUS END JOINING; KEGG MEDICUS REFERENCE V D J RECOMBINATION	27	catalytic activity, acting on DNA DNA binding	DNA recombination	nonhomologous end joining complex nucleoplasm	severe combined immunodeficiency
Carbohydrate Metabolism Disorder (glycosaminoglycan disorders, Defective CHST Genes)	REACTOME DEFECTIVE CHST14 CAUSES EDS MUSCULOCONTRACTURAL TYPE; REACTOME DEFECTIVE CHST3 CAUSES SEDCJD; REACTOME DEFECTIVE CHSY1 CAUSES TPBS; REACTOME DERMATAN SULFATE BIOSYNTHESIS; PID TOLL ENDOGENOUS PATHWAY; REACTOME DEFECTIVE B4GALT7 CAUSES EDS PROGEROID TYPE; REACTOME DISEASES ASSOCIATED WITH GLYCOSAMINOGLYCAN METABOLISM; REACTOME METABOLISM OF CARBOHYDRATES; REACTOME CS DS DEGRADATION; REACTOME A TETRASACCHARIDE LINKER SEQUENCE IS REQUIRED FOR GAG SYNTHESIS	319	catalytic activity structural constituent of nuclear pore	carbohydrate metabolic process	endomembrane system lysosomal lumen extracellular vesicle	Abnormal Urine Metabolite level
Immune Response to Viral Infections	REACTOME INFECTIOUS DISEASE; REACTOME HIV INFECTION; REACTOME HIV LIFE CYCLE; REACTOME VIRAL INFECTION PATHWAYS	1,019	structural molecule activity enzyme binding protein heterodimerization activity antigen binding	immune system process protein metabolic process	protein-containing complex extrinsic component of cytoplasmic side of plasma membrane	Abnormal cellular immune system morphology
Sphingolipid Metabolism	WP SPHINGOLIPID METABOLISM OVERVIEW; WP SPHINGOLIPID METABOLISM INTEGRATED PATHWAY; WP SPHINGOLIPID PATHWAY; WP BIOSYNTHESIS AND TURNOVER OF 1 DEOXY SPHINGOID BASES; KEGG SPHINGOLIPID METABOLISM	51	catalytic activity sphingolipid delta-4 denaturase activity	sphingolipid metabolic process	endomembrane system	cherry red spot of the macula foam cells decerebrate ridigity mental deterioration

Supplementary Table 3. *Enriched WMH ps-PRS in UKBB-II.*

Set	PRS.R2	Coefficient (Direction)	P	Num_SNP	Competitive.P
Extracellular Matrix	0.0074	1158	4.33E-14	3553	0.0001
Immune response to viral infections	0.0082	1530	2.02E-15	4766	0.0001
REACTOME INFECTIOUS DISEASE	0.0082	1530	2.02E-15	4766	0.0001
Metabolism of nucleotide sugars and glycolysis	0.0037	342	8.77E-08	510	0.0002
REACTOME HIV INFECTION	0.0047	544	2.03E-09	1043	0.0002
REACTOME HIV LIFE CYCLE	0.0036	381	1.81E-07	625	0.0002
WP GENES RELATED TO PRIMARY CILIUM DEVELOPMENT BASED ON CRISPR	0.0034	325	3.83E-07	538	0.0002
KEGG ALPHA LINOLENIC ACID METABOLISM	0.0030	113	1.50E-06	95	0.0003
REACTOME COENZYME A BIOSYNTHESIS	0.0031	49	1.23E-06	33	0.0003
REACTOME EXTRACELLULAR MATRIX ORGANIZATION	0.0054	829	1.11E-10	2605	0.0003
REACTOME VIRAL INFECTION PATHWAYS	0.0057	1100	4.00E-11	3425	0.0004
REACTOME DEFECTIVE CHST14 CAUSES EDS MUSCULOCONTRACTURAL TYPE	0.0022	47	3.87E-05	35	0.0008
REACTOME DEFECTIVE CHST3 CAUSES SEDCJD	0.0020	51	8.39E-05	40	0.0011
REACTOME SPERM MOTILITY AND TAXES	0.0020	72	9.38E-05	63	0.0011
WP VEGFA VEGFR2 SIGNALING	0.0049	935	8.98E-10	2927	0.0011
KEGG MEDICUS REFERENCE REGULATION OF GF RTK RAS ERK SIGNALING RAS UBIQUITINATION BY CUL3 COMPLEX	0.0017	41	2.75E-04	24	0.0012
REACTOME DEFECTIVE CHSY1 CAUSES TPBS	0.0019	51	1.24E-04	42	0.0012
KEGG MEDICUS REFERENCE CA2 PLCD ITPR SIGNALING PATHWAY	0.0022	99	4.13E-05	112	0.0013
REACTOME DERMATAN SULFATE BIOSYNTHESIS	0.0022	105	3.94E-05	104	0.0013
WP CONGENITAL GENERALIZED LIPODYSTROPHY	0.0023	104	2.52E-05	87	0.0013
KEGG MEDICUS REFERENCE MICROTUBULE DEPOLYMERIZATION	0.0020	102	8.82E-05	70	0.0015
REACTOME VITAMIN B5 PANTOTHENATE METABOLISM	0.0022	85	4.93E-05	92	0.0015
Organelle biosynthesis	0.0042	724	1.51E-08	2085	0.0017
BIOCARTA PRION PATHWAY	0.0020	99	1.10E-04	103	0.0019
WP SPHINGOLIPID METABOLISM OVERVIEW	0.0022	171	4.17E-05	208	0.0024
KEGG AMINO SUGAR AND NUCLEOTIDE SUGAR METABOLISM	0.0021	180	5.28E-05	271	0.0025
REACTOME ELASTIC FIBRE FORMATION	0.0021	193	5.60E-05	373	0.0028
REACTOME RESPONSE OF MTB TO PHAGOCYTOSIS	0.0017	110	2.62E-04	93	0.0029
WP OMEGA 3 OMEGA 6 FATTY ACID SYNTHESIS	0.0017	79	3.64E-04	90	0.0030

REACTOME MOLECULES ASSOCIATED WITH ELASTIC FIBRES	0.0021	179	6.07E-05	333	0.0031
KEGG GALACTOSE METABOLISM	0.0018	130	2.07E-04	161	0.0032
WP HEDGEHOG SIGNALING PATHWAY WP47	0.0017	79	3.19E-04	86	0.0032
Carbohydrate metabolism	0.0037	686	9.71E-08	1985	0.0033
PID TOLL ENDOGENOUS PATHWAY	0.0017	101	3.31E-04	115	0.0033
REACTOME DEFECTIVE B4GALT7 CAUSES EDS PROGEROID TYPE	0.0020	154	8.40E-05	234	0.0033
REACTOME MITOPHAGY	0.0016	100	4.55E-04	97	0.0034
REACTOME PLASMA LIPOPROTEIN REMODELING	0.0020	190	8.88E-05	279	0.0035
WP MALE STEROID HORMONES IN CARDIOMYOCYTE ENERGY METABOLISM	0.0016	83	4.90E-04	68	0.0036
WP SPHINGOLIPID METABOLISM INTEGRATED PATHWAY	0.0019	158	1.37E-04	204	0.0036
REACTOME GAP JUNCTION TRAFFICKING AND REGULATION	0.0019	156	1.65E-04	181	0.0038
REACTOME NEGATIVE REGULATION OF TCF DEPENDENT SIGNALING BY DVL INTERACTING PROTEINS	0.0013	36	1.33E-03	34	0.0038
WP ESTROGEN RECEPTOR PATHWAY	0.0016	78	5.31E-04	93	0.0038
REACTOME CDC42 GTPASE CYCLE	0.0033	508	4.91E-07	1626	0.0040
REACTOME SIGNALING BY VEGF	0.0028	385	4.38E-06	923	0.0040
Inflammatory signaling	0.0022	238	4.29E-05	490	0.0041
KEGG MEDICUS REFERENCE LPA GNA12 13 RHOA SIGNALING PATHWAY	0.0016	78	5.39E-04	77	0.0042
REACTOME PINK1 PRKN MEDIATED MITOPHAGY	0.0015	82	7.32E-04	73	0.0045
REACTOME ANCHORING OF THE BASAL BODY TO THE PLASMA MEMBRANE	0.0021	243	7.08E-05	475	0.0047
Signal Transduction	0.0053	1214	1.72E-10	4704	0.0048
REACTOME G0 AND EARLY G1	0.0015	109	7.85E-04	110	0.0050
REACTOME DISEASES ASSOCIATED WITH GLYCOSYLATION PRECURSOR BIOSYNTHESIS	0.0014	63	1.19E-03	63	0.0051
WP FAMILIAL HYPERLIPIDEMIA TYPE 3	0.0014	92	1.19E-03	91	0.0057
REACTOME DISEASES ASSOCIATED WITH GLYCOSAMINOGLYCAN METABOLISM	0.0018	188	2.22E-04	359	0.0059
WP EICOSANOID METABOLISM VIA CYCLOOXYGENASES COX	0.0017	135	3.50E-04	201	0.0059
REACTOME TRIGLYCERIDE BIOSYNTHESIS	0.0014	88	1.00E-03	82	0.0060
REACTOME METABOLISM OF CARBOHYDRATES	0.0032	625	6.50E-07	1894	0.0062
BIOCARTA SAM68 PATHWAY	0.0011	46	3.42E-03	27	0.0063
REACTOME CILIUM ASSEMBLY	0.0025	386	1.08E-05	954	0.0065
REACTOME MITOCHONDRIAL TRANSLATION	0.0017	190	3.28E-04	318	0.0069
WP MIRNA TARGETS IN ECM AND MEMBRANE RECEPTORS	0.0016	148	4.70E-04	268	0.0075
PID AVB3 INTEGRIN PATHWAY	0.0020	267	1.07E-04	639	0.0084
KEGG MEDICUS REFERENCE PROK PRKR GI ERK SIGNALING PATHWAY	0.0010	58	4.79E-03	46	0.0086

Non homologous end joining	0.0013	114	1.35E-03	117	0.0087
REACTOME CHYLOMICRON CLEARANCE	0.0010	49	4.91E-03	43	0.0091
KEGG MEDICUS VARIANT MUTATION INACTIVATED PROK2 TO PROK PRKR GI ERK SIGNALING PATHWAY	0.0010	53	4.89E-03	42	0.0096
REACTOME LAMININ INTERACTIONS	0.0016	163	4.28E-04	328	0.0096
Protein Localization and Fatty acid metabolism	0.0024	423	1.70E-05	1103	0.0098
WP SPHINGOLIPID PATHWAY	0.0015	161	6.23E-04	248	0.0099
REACTOME SIGNALING BY RECEPTOR TYROSINE KINASES	0.0046	1092	3.04E-09	4427	0.0100
PID EPHA FWDPATHWAY	0.0015	162	5.86E-04	341	0.0114
Sphingolipid metabolism	0.0015	178	6.37E-04	321	0.0119
REACTOME CALCITONIN LIKE LIGAND RECEPTORS	0.0010	45	6.77E-03	47	0.0119
KEGG MEDICUS REFERENCE PINK PARKIN MEDIATED AUTOPHAGOSOME FORMATION	0.0011	64	3.83E-03	80	0.0127
REACTOME CS DS DEGRADATION	0.0011	58	3.83E-03	80	0.0127
REACTOME SIGNALING BY MAPK MUTANTS	0.0009	47	7.56E-03	34	0.0129
REACTOME ORGANELLE BIOGENESIS AND MAINTENANCE	0.0024	450	1.85E-05	1355	0.0133
WP DISORDERS OF GALACTOSE METABOLISM	0.0011	61	4.21E-03	72	0.0135
REACTOME RHO GTPASE CYCLE	0.0039	893	3.90E-08	3690	0.0142
WP AEROBIC GLYCOLYSIS	0.0009	57	7.62E-03	52	0.0142
KEGG MEDICUS VARIANT MUTATION INACTIVATED OPTN TO PINK PARKIN MEDIATED AUTOPHAGOSOME FORMATION	0.0009	43	8.01E-03	43	0.0144
WP INTEGRIN MEDIATED CELL ADHESION	0.0020	314	8.74E-05	849	0.0146
KEGG NON HOMOLOGOUS END JOINING	0.0010	65	6.18E-03	59	0.0148
REACTOME FCGAMMA RECEPTOR FCGR DEPENDENT PHAGOCYTOSIS	0.0021	314	7.14E-05	893	0.0149
KEGG MEDICUS REFERENCE REGULATION OF GF RTK RAS ERK SIGNALING RASGAP	0.0009	41	9.51E-03	34	0.0156
WP BARDET BIEDL SYNDROME	0.0017	234	3.32E-04	598	0.0156
REACTOME A TETRASACCHARIDE LINKER SEQUENCE IS REQUIRED FOR GAG SYNTHESIS	0.0013	144	1.37E-03	303	0.0162
WP BIOSYNTHESIS AND TURNOVER OF 1 DEOXY SPHINGOID BASES	0.0010	80	6.45E-03	67	0.0175
BIOCARTA ACTINY PATHWAY	0.0009	49	8.43E-03	54	0.0176
REACTOME EPH EPHRIN SIGNALING	0.0019	292	1.44E-04	853	0.0189
REACTOME PRE NOTCH PROCESSING IN GOLGI	0.0010	64	6.36E-03	82	0.0196
KEGG SPHINGOLIPID METABOLISM	0.0012	137	2.53E-03	231	0.0198
REACTOME SIGNALING BY PDGF	0.0014	185	1.13E-03	430	0.0201
KEGG MEDICUS VARIANT MUTATION INACTIVATED P62 TO PINK PARKIN MEDIATED AUTOPHAGOSOME FORMATION	0.0009	49	1.02E-02	57	0.0205

PID AVB3 OPN PATHWAY	0.0012	132	2.16E-03	267	0.0206
REACTOME EARLY PHASE OF HIV LIFE CYCLE	0.0009	67	8.96E-03	61	0.0206
REACTOME TRANSCRIPTIONAL REGULATION BY E2F6	0.0010	89	5.86E-03	126	0.0206
REACTOME TRANSCRIPTION OF E2F TARGETS UNDER NEGATIVE CONTROL BY DREAM COMPLEX	0.0009	72	7.32E-03	71	0.0207
WP 2Q11 2 COPY NUMBER VARIATION SYNDROME	0.0009	57	9.34E-03	61	0.0211
BIOCARTA HIF PATHWAY	0.0009	56	8.74E-03	63	0.0213
REACTOME PEROXISOMAL LIPID METABOLISM	0.0011	100	4.60E-03	173	0.0217
REACTOME MITOCHONDRIAL RNA DEGRADATION	0.0008	61	1.11E-02	58	0.0230
KEGG SMALL CELL LUNG CANCER	0.0017	279	2.57E-04	792	0.0233
KEGG MEDICUS REFERENCE NON HOMOLOGOUS END JOINING	0.0009	67	9.17E-03	67	0.0235
REACTOME PEROXISOMAL PROTEIN IMPORT	0.0012	167	2.42E-03	315	0.0238
REACTOME PROTEIN LOCALIZATION	0.0016	275	4.58E-04	668	0.0239
REACTOME PEPTIDE HORMONE BIOSYNTHESIS	0.0008	40	1.61E-02	31	0.0250
REACTOME GABA SYNTHESIS RELEASE REUPTAKE AND DEGRADATION	0.0010	98	6.55E-03	167	0.0272
WP HEDGEHOG SIGNALING PATHWAY WP4249	0.0012	142	2.82E-03	318	0.0272
KEGG MEDICUS REFERENCE ACTIVATION OF PRC2.2 BY UBIQUITINATION OF H2AK119 IN GERMLINE GENES	0.0009	88	8.93E-03	126	0.0279
WP FOCAL ADHESION	0.0024	494	2.11E-05	1802	0.0293
REACTOME VITAMINS	0.0007	33	2.15E-02	19	0.0298
REACTOME MRNA DECAY BY 5 TO 3 EXORIBONUCLEASE	0.0008	61	1.65E-02	58	0.0303
REACTOME PARASITE INFECTION	0.0014	202	1.19E-03	553	0.0308
REACTOME TCF DEPENDENT SIGNALING IN RESPONSE TO WNT	0.0017	323	2.85E-04	969	0.0316
REACTOME ASSEMBLY OF COLLAGEN FIBRILS AND OTHER MULTIMERIC STRUCTURES	0.0013	202	1.35E-03	543	0.0318
WP RESISTIN AS A REGULATOR OF INFLAMMATION	0.0011	148	3.13E-03	354	0.0319
REACTOME EPIGENETIC REGULATION OF GENE EXPRESSION	0.0014	231	1.18E-03	575	0.0321
REACTOME ALPHA LINOLENIC OMEGA3 AND LINOLEIC OMEGA6 ACID METABOLISM	0.0008	43	1.45E-02	69	0.0322
KEGG MEDICUS REFERENCE V D J RECOMBINATION	0.0007	59	1.78E-02	60	0.0329
REACTOME MHC CLASS II ANTIGEN PRESENTATION	0.0013	217	1.78E-03	513	0.0348
REACTOME NEGATIVE REGULATION OF MAPK PATHWAY	0.0010	115	6.94E-03	205	0.0354
KEGG PEROXISOME	0.0011	171	3.26E-03	387	0.0356

WP EICOSANOID METABOLISM VIA LIPOOXYGENASES LOX	0.0008	73	1.26E-02	137	0.0356
REACTOME EPHA MEDIATED GROWTH CONE COLLAPSE	0.0010	121	4.86E-03	271	0.0357
WP PARKIN UBIQUITIN PROTEASOMAL SYSTEM PATHWAY	0.0010	125	6.31E-03	219	0.0360
REACTOME RUNX2 REGULATES OSTEOBLAST DIFFERENTIATION	0.0009	91	9.40E-03	181	0.0372
REACTOME RESPIRATORY ELECTRON TRANSPORT	0.0011	171	3.65E-03	374	0.0380
KEGG MEDICUS ENV FACTOR IRON TO ANTEROGRADE AXONAL TRANSPORT	0.0007	43	2.52E-02	40	0.0381
WP ONE CARBON METABOLISM	0.0009	106	8.20E-03	201	0.0385
WP GLYCOLYSIS AND GLUCONEOGENESIS	0.0008	103	1.08E-02	173	0.0392
BIOCARTA PARKIN PATHWAY	0.0006	28	2.86E-02	35	0.0416
REACTOME REGULATION OF HMOX1 EXPRESSION AND ACTIVITY	0.0006	34	3.05E-02	34	0.0418
BIOCARTA NPP1 PATHWAY	0.0008	59	1.65E-02	121	0.0423
PID BARD1 PATHWAY	0.0008	93	1.46E-02	138	0.0423
KEGG UBIQUITIN MEDIATED PROTEOLYSIS	0.0013	227	1.92E-03	595	0.0438
REACTOME SIGNALING BY RHO GTPASES MIRO GTPASES AND RHOBTB3	0.0039	1003	5.42E-08	4685	0.0455
PID_EPHA2_FWD_PATHWAY	0.0008	92	1.41E-02	181	0.0486

Supplementary Table 4. *3C-Dijon ps-PRS enrichment. All ps-PRS enriched in UKBB.*

Set	PRS.R2	Coefficient (Direction)	P	Num_SNP	Competitive.P
WP ESTROGEN RECEPTOR PATHWAY	0.012	242	3.76E-05	94	0.0001
REACTOME G0 AND EARLY G1	0.008	278	5.05E-04	109	0.0009
REACTOME TRANSCRIPTION OF E2F TARGETS UNDER NEGATIVE CONTROL BY DREAM COMPLEX	0.008	227	6.21E-04	71	0.0009
REACTOME ALPHA LINOLENIC OMEGA3 AND LINOLEIC OMEGA6 ACID METABOLISM	0.007	138	2.07E-03	68	0.0034
REACTOME MITOCHONDRIAL TRANSLATION	0.007	415	2.11E-03	316	0.0036
REACTOME SIGNALING BY PDGF	0.006	396	4.88E-03	422	0.0082
Organelle biosynthesis	0.006	977	2.66E-03	2051	0.0163
REACTOME ORGANELLE BIOGENESIS AND MAINTENANCE	0.005	738	5.62E-03	1337	0.0163
REACTOME ELASTIC FIBRE FORMATION	0.004	299	1.38E-02	369	0.0219
REACTOME DERMATAN SULFATE BIOSYNTHESIS	0.004	-150	1.81E-02	103	0.0229
REACTOME DEFECTIVE CHST3 CAUSES SEDCJD	0.004	-73	2.39E-02	40	0.0255
REACTOME MOLECULES ASSOCIATED WITH ELASTIC FIBRES	0.004	261	2.15E-02	329	0.0318
KEGG MEDICUS REFERENCE MICROTUBULE DEPOLYMERIZATION	0.003	142	2.99E-02	68	0.0362
PID TOLL ENDOGENOUS PATHWAY	0.003	-146	3.16E-02	110	0.0408
REACTOME ANCHORING OF THE BASAL BODY TO THE PLASMA MEMBRANE	0.003	329	2.97E-02	463	0.0415
REACTOME DEFECTIVE CHSY1 CAUSES TPBS	0.003	-65	4.28E-02	41	0.0446
REACTOME DEFECTIVE CHST14 CAUSES EDS MUSCULOCONTRACTURAL TYPE	0.003	-57	4.37E-02	35	0.0462
WP HEDGEHOG SIGNALING PATHWAY WP4249	0.003	251	3.54E-02	310	0.0478
REACTOME CALCITONIN LIKE LIGAND RECEPTORS	0.003	85	4.52E-02	47	0.0494
KEGG MEDICUS VARIANT MUTATION INACTIVATED OPTN TO PINK PARKIN MEDIATED AUTOPHAGOSOME FORMATION	0.002	74	6.70E-02	42	0.0714
KEGG MEDICUS ENV FACTOR IRON TO ANTEROGRADE AXONAL TRANSPORT	0.002	86	6.74E-02	43	0.0717
REACTOME PLASMA LIPOPROTEIN REMODELING	0.002	223	7.24E-02	279	0.0878
REACTOME NEGATIVE REGULATION OF MAPK PATHWAY	0.002	178	8.80E-02	201	0.1010
Extracellular Matrix	0.004	888	2.15E-02	3515	0.1092
WP MIRNA TARGETS IN ECM AND MEMBRANE RECEPTORS	0.002	171	1.09E-01	265	0.1264

WP HEDGEHOG SIGNALING PATHWAY WP47	0.002	86	1.25E-01	84	0.1350
REACTOME CS DS DEGRADATION	0.002	-74	1.28E-01	78	0.1374
REACTOME ASSEMBLY OF COLLAGEN FIBRILS AND OTHER MULTIMERIC STRUCTURES	0.002	253	1.09E-01	533	0.1383
WP OMEGA 3 OMEGA 6 FATTY ACID SYNTHESIS	0.002	81	1.39E-01	89	0.1474
WP PARKIN UBIQUITIN PROTEASOMAL SYSTEM PATHWAY	0.002	-173	1.42E-01	216	0.1594
REACTOME CILIUM ASSEMBLY	0.002	361	1.10E-01	942	0.1624
WP AEROBIC GLYCOLYSIS	0.001	-79	1.70E-01	52	0.1751
REACTOME HIV LIFE CYCLE	0.001	277	1.43E-01	618	0.1819
KEGG MEDICUS REFERENCE CA2 PLCD ITPR SIGNALING PATHWAY	0.001	-82	1.79E-01	111	0.1900
REACTOME VIRAL INFECTION PATHWAYS	0.003	821	5.37E-02	3380	0.1931
PID AVB3 INTEGRIN PATHWAY	0.001	240	1.60E-01	628	0.1991
REACTOME EXTRACELLULAR MATRIX ORGANIZATION	0.002	556	8.51E-02	2572	0.2115
PID EPHA2 FWD PATHWAY	0.001	122	2.04E-01	182	0.2207
REACTOME LAMININ INTERACTIONS	0.001	150	1.96E-01	325	0.2207
WP BARDET BIEDL SYNDROME	0.001	214	1.99E-01	586	0.2377
REACTOME GAP JUNCTION TRAFFICKING AND REGULATION	0.001	132	2.24E-01	179	0.2424
WP EICOSANOID METABOLISM VIA CYCLOOXYGENASES COX	0.001	113	2.30E-01	200	0.2468
REACTOME TCF DEPENDENT SIGNALING IN RESPONSE TO WNT	0.001	300	1.87E-01	957	0.2470
REACTOME MITOCHONDRIAL RNA DEGRADATION	0.001	78	2.46E-01	57	0.2549
REACTOME PARASITE INFECTION	0.001	-194	2.20E-01	553	0.2569
WP EICOSANOID METABOLISM VIA LIPOOXYGENASES LOX	0.001	85	2.44E-01	137	0.2579
REACTOME VITAMIN B5 PANTOTHENATE METABOLISM	0.001	60	2.52E-01	92	0.2612
KEGG MEDICUS REFERENCE REGULATION OF GF RTK RAS ERK SIGNALING RASGAP	0.001	43	2.57E-01	33	0.2650
REACTOME PEROXISOMAL LIPID METABOLISM	0.001	102	2.48E-01	172	0.2674
REACTOME PEROXISOMAL PROTEIN IMPORT	0.001	158	2.48E-01	312	0.2713
WP CONGENITAL GENERALIZED LIPODYSTROPHY	0.001	67	2.69E-01	86	0.2782
Immune response to viral infections	0.002	886	7.71E-02	4712	0.2936
WP ONE CARBON METABOLISM	0.001	108	2.77E-01	202	0.2938
REACTOME RESPONSE OF MTB TO PHAGOCYTOSIS	0.001	81	2.93E-01	92	0.3016
REACTOME INFECTIOUS DISEASE	0.002	886	7.71E-02	4712	0.3024
PID EPHA FWDPATHWAY	0.001	132	2.86E-01	340	0.3061

WP BIOSYNTHESIS AND TURNOVER OF 1 DEOXY SPHINGOID BASES	0.001	-72	3.16E-01	68	0.3218
KEGG MEDICUS REFERENCE REGULATION OF GF RTK RAS ERK SIGNALING RAS UBIQUITINATION BY CUL3 COMPLEX	0.001	26	3.21E-01	23	0.3243
REACTOME SPERM MOTILITY AND TAXES	0.001	47	3.40E-01	62	0.3440
BIOCARTA HIF PATHWAY	0.001	50	3.44E-01	63	0.3473
REACTOME TRIGLYCERIDE BIOSYNTHESIS	0.001	-64	3.45E-01	81	0.3539
REACTOME FCGAMMA RECEPTOR FCGR DEPENDENT PHAGOCYTOSIS	0.001	-210	2.99E-01	887	0.3587
WP DISORDERS OF GALACTOSE METABOLISM	0.001	49	3.50E-01	72	0.3606
WP GENES RELATED TO PRIMARY CILIUM DEVELOPMENT BASED ON CRISPR	0.001	149	3.55E-01	529	0.3914
Signal Transduction	0.002	727	1.34E-01	4676	0.3918
BIOCARTA NPP1 PATHWAY	0.001	52	3.87E-01	119	0.3979
REACTOME A TETRASACCHARIDE LINKER SEQUENCE IS REQUIRED FOR GAG SYNTHESIS	0.001	-96	3.87E-01	300	0.4076
BIOCARTA PARKIN PATHWAY	0.000	-26	4.10E-01	32	0.4167
Sphingolipid metabolism	0.000	110	4.04E-01	319	0.4316
REACTOME EPH EPHRIN SIGNALING	0.001	178	3.79E-01	848	0.4353
REACTOME SIGNALING BY RECEPTOR TYROSINE KINASES	0.001	644	1.72E-01	4400	0.4428
REACTOME VITAMINS	0.000	27	4.49E-01	19	0.4486
WP FOCAL ADHESION	0.001	290	3.30E-01	1787	0.4497
BIOCARTA PRION PATHWAY	0.000	51	4.38E-01	105	0.4515
REACTOME DEFECTIVE B4GALT7 CAUSES EDS PROGEROID TYPE	0.000	-77	4.38E-01	233	0.4564
WP GLYCOLYSIS AND GLUCONEOGENESIS	0.000	82	4.55E-01	173	0.4679
Carbohydrate metabolism	0.001	-292	3.57E-01	1963	0.4939
REACTOME NEGATIVE REGULATION OF TCF DEPENDENT SIGNALING BY DVL INTERACTING PROTEINS	0.000	19	5.01E-01	34	0.5061
KEGG MEDICUS REFERENCE PINK PARKIN MEDIATED AUTOPHAGOSOME FORMATION	0.000	36	5.16E-01	78	0.5269
Inflammatory signaling	0.000	96	5.03E-01	487	0.5338
REACTOME METABOLISM OF CARBOHYDRATES	0.000	-255	4.08E-01	1877	0.5350
WP INTEGRIN MEDIATED CELL ADHESION	0.000	138	4.90E-01	840	0.5405
REACTOME RESPIRATORY ELECTRON TRANSPORT	0.000	95	5.20E-01	369	0.5447
KEGG GALACTOSE METABOLISM	0.000	56	5.39E-01	158	0.5468

REACTOME COENZYME A BIOSYNTHESIS	0.000	15	5.57E-01	33	0.5575
KEGG SMALL CELL LUNG CANCER	0.000	118	5.26E-01	788	0.5660
BIOCARTA SAM68 PATHWAY	0.000	-22	5.71E-01	27	0.5731
Metabolism of nucleotide sugars and glycolysis	0.000	96	5.60E-01	504	0.5874
Non homologous end joining	0.000	48	5.76E-01	114	0.5874
REACTOME DISEASES ASSOCIATED WITH GLYCOSYLATION PRECURSOR BIOSYNTHESIS	0.000	26	5.85E-01	63	0.5878
KEGG AMINO SUGAR AND NUCLEOTIDE SUGAR METABOLISM	0.000	-66	5.77E-01	269	0.5892
KEGG MEDICUS VARIANT MUTATION INACTIVATED P62 TO PINK PARKIN MEDIATED AUTOPHAGOSOME FORMATION	0.000	25	5.92E-01	56	0.5972
PID AVB3 OPN PATHWAY	0.000	-59	5.93E-01	264	0.6037
WP RESISTIN AS A REGULATOR OF INFLAMMATION	0.000	67	5.92E-01	352	0.6161
REACTOME SIGNALING BY MAPK MUTANTS	0.000	-22	6.23E-01	33	0.6248
REACTOME REGULATION OF HMOX1 EXPRESSION AND ACTIVITY	0.000	-19	6.33E-01	34	0.6349
REACTOME PROTEIN LOCALIZATION	0.000	103	6.11E-01	662	0.6366
REACTOME HIV INFECTION	0.000	126	5.91E-01	1032	0.6476
KEGG MEDICUS REFERENCE ACTIVATION OF PRC2.2 BY UBIQUITINATION OF H2AK119 IN GERMLINE GENES	0.000	-38	6.62E-01	135	0.6667
REACTOME RUNX2 REGULATES OSTEOBLAST DIFFERENTIATION	0.000	39	6.67E-01	179	0.6681
KEGG PEROXISOME	0.000	64	6.50E-01	384	0.6771
REACTOME PRE NOTCH PROCESSING IN GOLGI	0.000	-24	6.87E-01	82	0.6899
WP VEGFA VEGFR2 SIGNALING	0.000	239	5.42E-01	2898	0.6977
KEGG ALPHA LINOLENIC ACID METABOLISM	0.000	21	7.16E-01	94	0.7196
REACTOME CHYLOMICRON CLEARANCE	0.000	17	7.16E-01	43	0.7276
KEGG MEDICUS REFERENCE NON HOMOLOGOUS END JOINING	0.000	-21	7.33E-01	66	0.7372
WP 2Q11 2 COPY NUMBER VARIATION SYNDROME	0.000	-17	7.50E-01	59	0.7508
REACTOME PEPTIDE HORMONE BIOSYNTHESIS	0.000	14	7.52E-01	29	0.7574
PID BARD1 PATHWAY	0.000	27	7.74E-01	136	0.7784
REACTOME EARLY PHASE OF HIV LIFE CYCLE	0.000	17	7.90E-01	60	0.7871
KEGG NON HOMOLOGOUS END JOINING	0.000	14	8.11E-01	58	0.8161
WP SPHINGOLIPID METABOLISM INTEGRATED PATHWAY	0.000	-24	8.14E-01	204	0.8195

KEGG MEDICUS VARIANT MUTATION INACTIVATED PROK2 TO PROK PRKR GI ERK SIGNALING PATHWAY	0.000	10	8.31E-01	42	0.8349
REACTOME MRNA DECAY BY 5 TO 3 EXORIBONUCLEASE	0.000	-14	8.33E-01	58	0.8385
WP SPHINGOLIPID PATHWAY	0.000	23	8.45E-01	247	0.8491
WP FAMILIAL HYPERLIPIDEMIA TYPE 3	0.000	-14	8.50E-01	93	0.8515
REACTOME GABA SYNTHESIS RELEASE REUPTAKE AND DEGRADATION	0.000	-17	8.56E-01	169	0.8609
KEGG UBIQUITIN MEDIATED PROTEOLYSIS	0.000	-36	8.46E-01	584	0.8610
Protein Localization and Fatty acid metabolism	0.000	53	8.33E-01	1092	0.8631
REACTOME CDC42 GTPASE CYCLE	0.000	-58	8.25E-01	1606	0.8656
REACTOME EPHA MEDIATED GROWTH CONE COLLAPSE	0.000	-16	8.81E-01	269	0.8872
REACTOME SIGNALING BY RHO GTPASES MIRO GTPASES AND RHOBTB3	0.000	145	7.65E-01	4631	0.8876
REACTOME SIGNALING BY VEGF	0.000	-32	8.79E-01	919	0.8947
KEGG MEDICUS REFERENCE V D J RECOMBINATION	0.000	-8	8.96E-01	58	0.8955
WP SPHINGOLIPID METABOLISM OVERVIEW	0.000	14	8.96E-01	208	0.8988
REACTOME TRANSCRIPTIONAL REGULATION BY E2F6	0.000	-9	9.14E-01	135	0.9180
KEGG MEDICUS REFERENCE PROK PRKR GI ERK SIGNALING PATHWAY	0.000	5	9.21E-01	46	0.9211
REACTOME RHO GTPASE CYCLE	0.000	75	8.55E-01	3646	0.9218
REACTOME MHC CLASS II ANTIGEN PRESENTATION	0.000	18	9.21E-01	509	0.9247
REACTOME MITOPHAGY	0.000	6	9.34E-01	97	0.9323
WP MALE STEROID HORMONES IN CARDIOMYOCYTE ENERGY METABOLISM	0.000	-5	9.34E-01	65	0.9333
REACTOME DISEASES ASSOCIATED WITH GLYCOSAMINOGLYCAN METABOLISM	0.000	10	9.38E-01	350	0.9410
REACTOME PINK1 PRKN MEDIATED MITOPHAGY	0.000	5	9.39E-01	73	0.9416
BIOCARTA ACTINY PATHWAY	0.000	3	9.43E-01	54	0.9435
REACTOME EPIGENETIC REGULATION OF GENE EXPRESSION	0.000	4	9.81E-01	571	0.9812
KEGG SPHINGOLIPID METABOLISM	0.000	-2	9.87E-01	230	0.9886
KEGG MEDICUS REFERENCE LPA GNA12 13 RHOA SIGNALING PATHWAY	0.000	-1	9.93E-01	77	0.9928

Supplementary Table 5. *i-Share ps-PRS enrichment. All ps-PRS enriched in UKBB.*

Set	PRS.R2	Coefficient (Direction)	P	Num_SNP	Competitive.P
REACTOME ANCHORING OF THE BASAL BODY TO THE PLASMA MEMBRANE	0.0051	438	0.0047	445	0.0056
REACTOME TCF DEPENDENT SIGNALING IN RESPONSE TO WNT	0.0045	618	0.0081	899	0.0085
WP FOCAL ADHESION	0.0041	777	0.0113	1688	0.0099
REACTOME VITAMIN B5 PANTOTHENATE METABOLISM	0.0039	135	0.0136	91	0.0157
KEGG MEDICUS REFERENCE ACTIVATION OF PRC2.2 BY UBIQUITINATION OF H2AK119 IN GERMLINE GENES	0.0035	220	0.0193	129	0.0190
Protein Localization and Fatty acid metabolism	0.0034	611	0.0200	1037	0.0192
Signal Transduction	0.0029	1118	0.0322	4438	0.0236
REACTOME DISEASES ASSOCIATED WITH GLYCOSAMINOGLYCAN METABOLISM	0.0032	295	0.0242	343	0.0247
REACTOME ALPHA LINOLENIC OMEGA3 AND LINOLEIC OMEGA6 ACID METABOLISM	0.0032	104	0.0243	64	0.0274
BIOCARTA NPP1 PATHWAY	0.0032	147	0.0257	114	0.0276
PID EPHA2 FWD PATHWAY	0.0031	216	0.0267	173	0.0282
KEGG SMALL CELL LUNG CANCER	0.0030	460	0.0304	747	0.0313
WP RESISTIN AS A REGULATOR OF INFLAMMATION	0.0029	286	0.0330	338	0.0320
WP OMEGA 3 OMEGA 6 FATTY ACID SYNTHESIS	0.0028	129	0.0346	85	0.0356
PID TOLL ENDOGENOUS PATHWAY	0.0028	154	0.0346	104	0.0359
REACTOME SIGNALING BY RECEPTOR TYROSINE KINASES	0.0025	997	0.0478	4174	0.0417
Carbohydrate metabolism	0.0024	661	0.0539	1856	0.0471
REACTOME DEFECTIVE CHSY1 CAUSES TPBS	0.0025	62	0.0497	38	0.0505
REACTOME DEFECTIVE CHST14 CAUSES EDS MUSCULOCONTRACTURAL TYPE	0.0025	55	0.0476	32	0.0519
REACTOME DEFECTIVE CHST3 CAUSES SEDCJD	0.0023	60	0.0555	37	0.0564
BIOCARTA PARKIN PATHWAY	0.0023	60	0.0551	30	0.0571
BIOCARTA ACTINY PATHWAY	0.0023	97	0.0582	54	0.0581
REACTOME NEGATIVE REGULATION OF MAPK PATHWAY	0.0023	209	0.0568	192	0.0597
KEGG ALPHA LINOLENIC ACID METABOLISM	0.0023	116	0.0568	88	0.0617
REACTOME DEFECTIVE B4GALT7 CAUSES EDS PROGEROID TYPE	0.0023	188	0.0592	227	0.0647
BIOCARTA HIF PATHWAY	0.0022	95	0.0644	58	0.0658
KEGG PEROXISOME	0.0022	283	0.0650	364	0.0659

REACTOME METABOLISM OF CARBOHYDRATES	0.0020	596	0.0735	1776	0.0682
REACTOME PROTEIN LOCALIZATION	0.0021	379	0.0665	630	0.0696
WP ESTROGEN RECEPTOR PATHWAY	0.0022	108	0.0657	88	0.0699
WP EICOSANOID METABOLISM VIA CYCLOOXYGENASES COX	0.0021	181	0.0718	188	0.0724
REACTOME CILIUM ASSEMBLY	0.0019	404	0.0807	910	0.0817
REACTOME TRANSCRIPTIONAL REGULATION BY E2F6	0.0019	156	0.0840	128	0.0879
REACTOME COENZYME A BIOSYNTHESIS	0.0019	45	0.0876	33	0.0891
REACTOME DERMATAN SULFATE BIOSYNTHESIS	0.0018	109	0.0886	95	0.0911
Inflammatory signaling	0.0017	247	0.1070	470	0.1069
WP 2Q11 2 COPY NUMBER VARIATION SYNDROME	0.0016	90	0.1113	56	0.1179
REACTOME ORGANELLE BIOGENESIS AND MAINTENANCE	0.0015	426	0.1217	1286	0.1194
REACTOME SIGNALING BY PDGF	0.0015	227	0.1193	401	0.1246
REACTOME A TETRASACCHARIDE LINKER SEQUENCE IS REQUIRED FOR GAG SYNTHESIS	0.0015	177	0.1278	291	0.1305
REACTOME PEROXISOMAL PROTEIN IMPORT	0.0015	220	0.1313	295	0.1307
REACTOME EARLY PHASE OF HIV LIFE CYCLE	0.0015	-100	0.1259	57	0.1317
WP HEDGEHOG SIGNALING PATHWAY WP4249	0.0014	-184	0.1373	287	0.1367
WP ONE CARBON METABOLISM	0.0013	153	0.1534	193	0.1524
KEGG MEDICUS REFERENCE LPA GNA12 13 RHOA SIGNALING PATHWAY	0.0012	-81	0.1688	76	0.1707
REACTOME CALCITONIN LIKE LIGAND RECEPTORS	0.0011	65	0.1811	46	0.1872
WP AEROBIC GLYCOLYSIS	0.0010	73	0.2034	50	0.2046
WP BARDET BIEDL SYNDROME	0.0010	209	0.2100	565	0.2115
Organelle biosynthesis	0.0009	402	0.2290	1973	0.2219
REACTOME SPERM MOTILITY AND TAXES	0.0009	-60	0.2243	59	0.2271
REACTOME CS DS DEGRADATION	0.0009	57	0.2313	72	0.2279
REACTOME SIGNALING BY MAPK MUTANTS	0.0009	50	0.2346	30	0.2418
REACTOME MITOPHAGY	0.0008	84	0.2521	95	0.2555
REACTOME RESPONSE OF MTB TO PHAGOCYTOSIS	0.0008	85	0.2591	88	0.2605
WP EICOSANOID METABOLISM VIA LIPOOXYGENASES LOX	0.0007	85	0.2806	133	0.2800
REACTOME LAMININ INTERACTIONS	0.0007	128	0.2915	307	0.2940
WP GLYCOLYSIS AND GLUCONEOGENESIS	0.0006	113	0.3164	170	0.3156
WP MALE STEROID HORMONES IN CARDIOMYOCYTE ENERGY METABOLISM	0.0006	60	0.3187	65	0.3209

WP INTEGRIN MEDIATED CELL ADHESION	0.0006	203	0.3352	806	0.3306
KEGG MEDICUS REFERENCE REGULATION OF GF RTK RAS ERK SIGNALING RAS UBIQUITINATION BY CUL3 COMPLEX	0.0006	27	0.3273	23	0.3333
REACTOME PEROXISOMAL LIPID METABOLISM	0.0005	82	0.3675	164	0.3670
KEGG SPHINGOLIPID METABOLISM	0.0005	102	0.3855	218	0.3768
REACTOME VITAMINS	0.0005	31	0.3809	19	0.3802
WP DISORDERS OF GALACTOSE METABOLISM	0.0005	46	0.3865	67	0.3834
KEGG MEDICUS REFERENCE REGULATION OF GF RTK RAS ERK SIGNALING RASGAP	0.0004	35	0.4117	34	0.4195
WP BIOSYNTHESIS AND TURNOVER OF 1 DEOXY SPHINGOID BASES	0.0004	58	0.4336	63	0.4326
KEGG GALACTOSE METABOLISM	0.0004	70	0.4436	149	0.4380
WP GENES RELATED TO PRIMARY CILIUM DEVELOPMENT BASED ON CRISPR	0.0004	123	0.4407	511	0.4430
WP SPHINGOLIPID METABOLISM OVERVIEW	0.0004	82	0.4520	199	0.4472
Non homologous end joining	0.0004	-69	0.4422	110	0.4515
Extracellular Matrix	0.0003	285	0.4937	3341	0.4742
REACTOME FCGAMMA RECEPTOR FCGR DEPENDENT PHAGOCYTOSIS	0.0003	147	0.4850	852	0.4825
WP MIRNA TARGETS IN ECM AND MEMBRANE RECEPTORS	0.0003	77	0.4865	252	0.4918
REACTOME DISEASES ASSOCIATED WITH GLYCOSYLATION PRECURSOR BIOSYNTHESIS	0.0003	35	0.4889	61	0.4932
REACTOME EXTRACELLULAR MATRIX ORGANIZATION	0.0003	-227	0.5031	2448	0.4957
REACTOME PEPTIDE HORMONE BIOSYNTHESIS	0.0003	28	0.4985	28	0.5060
KEGG MEDICUS ENV FACTOR IRON TO ANTEROGRADE AXONAL TRANSPORT	0.0003	-30	0.5200	41	0.5203
REACTOME EPH EPHRIN SIGNALING	0.0002	128	0.5386	814	0.5387
REACTOME HIV LIFE CYCLE	0.0002	-114	0.5547	589	0.5555
WP VEGFA VEGFR2 SIGNALING	0.0002	233	0.5664	2740	0.5568
WP CONGENITAL GENERALIZED LIPODYSTROPHY	0.0002	36	0.5596	82	0.5591
REACTOME PLASMA LIPOPROTEIN REMODELING	0.0002	75	0.5650	267	0.5696
KEGG MEDICUS VARIANT MUTATION INACTIVATED OPTN TO PINK PARKIN MEDIATED AUTOPHAGOSOME FORMATION	0.0002	24	0.5710	43	0.5752
REACTOME MHC CLASS II ANTIGEN PRESENTATION	0.0002	99	0.5757	483	0.5758
REACTOME EPIGENETIC REGULATION OF GENE EXPRESSION	0.0002	105	0.5714	546	0.5774

REACTOME SIGNALING BY RHO GTPASES MIRO GTPASES AND RHOBTB3	0.0002	267	0.5849	4419	0.5781
WP FAMILIAL HYPERLIPIDEMIA TYPE 3	0.0002	42	0.5839	90	0.5818
REACTOME ELASTIC FIBRE FORMATION	0.0002	-68	0.5813	355	0.5821
REACTOME GABA SYNTHESIS RELEASE REUPTAKE AND DEGRADATION	0.0002	51	0.5898	164	0.5873
REACTOME TRIGLYCERIDE BIOSYNTHESIS	0.0002	35	0.6014	77	0.6030
REACTOME EPHA MEDIATED GROWTH CONE COLLAPSE	0.0002	59	0.6072	258	0.6079
REACTOME G0 AND EARLY G1	0.0002	44	0.6118	105	0.6083
WP SPHINGOLIPID METABOLISM INTEGRATED PATHWAY	0.0002	54	0.6159	195	0.6100
REACTOME PRE NOTCH PROCESSING IN GOLGI	0.0001	-29	0.6300	80	0.6344
REACTOME REGULATION OF HMOX1 EXPRESSION AND ACTIVITY	0.0001	-19	0.6432	33	0.6518
PID EPHA FWDPATHWAY	0.0001	55	0.6720	327	0.6684
Metabolism of nucleotide sugars and glycolysis	0.0001	71	0.6728	480	0.6691
REACTOME NEGATIVE REGULATION OF TCF DEPENDENT SIGNALING BY DVL INTERACTING PROTEINS	0.0001	13	0.6627	33	0.6714
BIOCARTA SAM68 PATHWAY	0.0001	18	0.6790	27	0.6882
PID AVB3 INTEGRIN PATHWAY	0.0001	69	0.6916	585	0.6990
REACTOME MRNA DECAY BY 5 TO 3 EXORIBONUCLEASE	0.0001	-27	0.7025	56	0.7005
REACTOME MITOCHONDRIAL RNA DEGRADATION	0.0001	25	0.6996	53	0.7007
REACTOME RESPIRATORY ELECTRON TRANSPORT	0.0001	58	0.7050	358	0.7061
REACTOME CHYLOMICRON CLEARANCE	0.0001	-17	0.7162	40	0.7141
KEGG NON HOMOLOGOUS END JOINING	0.0001	21	0.7211	55	0.7231
KEGG MEDICUS REFERENCE CA2 PLCD ITPR SIGNALING PATHWAY	0.0001	-22	0.7330	107	0.7316
REACTOME HIV INFECTION	0.0001	81	0.7429	976	0.7407
Sphingolipid metabolism	0.0001	43	0.7503	303	0.7465
KEGG MEDICUS REFERENCE NON HOMOLOGOUS END JOINING	0.0001	-21	0.7500	63	0.7470
KEGG MEDICUS VARIANT MUTATION INACTIVATED PROK2 TO PROK PRKR GI ERK SIGNALING PATHWAY	0.0001	16	0.7553	41	0.7535
KEGG MEDICUS VARIANT MUTATION INACTIVATED P62 TO PINK PARKIN MEDIATED AUTOPHAGOSOME FORMATION	0.0001	15	0.7562	56	0.7544
REACTOME TRANSCRIPTION OF E2F TARGETS UNDER NEGATIVE CONTROL BY DREAM COMPLEX	0.0000	19	0.7902	68	0.7937

REACTOME GAP JUNCTION TRAFFICKING AND REGULATION	0.0000	-28	0.7958	174	0.7942
REACTOME MOLECULES ASSOCIATED WITH ELASTIC FIBRES	0.0000	-28	0.8045	316	0.7981
KEGG MEDICUS REFERENCE V D J RECOMBINATION	0.0000	-16	0.8044	57	0.8039
REACTOME RUNX2 REGULATES OSTEOBLAST DIFFERENTIATION	0.0000	22	0.8094	176	0.8058
REACTOME RHO GTPASE CYCLE	0.0000	96	0.8194	3480	0.8131
PID BARD1 PATHWAY	0.0000	21	0.8234	129	0.8246
PID AVB3 OPN PATHWAY	0.0000	25	0.8305	252	0.8295
REACTOME ASSEMBLY OF COLLAGEN FIBRILS AND OTHER MULTIMERIC STRUCTURES	0.0000	-34	0.8394	501	0.8346
KEGG MEDICUS REFERENCE PINK PARKIN MEDIATED AUTOPHAGOSOME FORMATION	0.0000	12	0.8360	77	0.8380
WP HEDGEHOG SIGNALING PATHWAY WP47	0.0000	11	0.8456	82	0.8407
REACTOME CDC42 GTPASE CYCLE	0.0000	48	0.8585	1535	0.8574
REACTOME SIGNALING BY VEGF	0.0000	-39	0.8613	863	0.8617
KEGG MEDICUS REFERENCE PROK PRKR GI ERK SIGNALING PATHWAY	0.0000	9	0.8633	45	0.8628
REACTOME MITOCHONDRIAL TRANSLATION	0.0000	-21	0.8825	310	0.8774
REACTOME PARASITE INFECTION	0.0000	22	0.8928	527	0.8908
BIOCARTA PRION PATHWAY	0.0000	-9	0.8918	92	0.8912
WP SPHINGOLIPID PATHWAY	0.0000	16	0.8919	233	0.8925
KEGG MEDICUS REFERENCE MICROTUBULE DEPOLYMERIZATION	0.0000	8	0.9085	62	0.9048
REACTOME PINK1 PRKN MEDIATED MITOPHAGY	0.0000	7	0.9092	72	0.9086
KEGG UBIQUITIN MEDIATED PROTEOLYSIS	0.0000	-20	0.9162	562	0.9173
REACTOME VIRAL INFECTION PATHWAYS	0.0000	32	0.9424	3227	0.9390
WP PARKIN UBIQUITIN PROTEASOMAL SYSTEM PATHWAY	0.0000	9	0.9400	206	0.9428
KEGG AMINO SUGAR AND NUCLEOTIDE SUGAR METABOLISM	0.0000	7	0.9530	253	0.9527
REACTOME INFECTIOUS DISEASE	0.0000	-8	0.9870	4493	0.9872
Immune response to viral infections	0.0000	-8	0.9870	4493	0.9882

Supplementary Table 6. *Memento WMH ps-PRS enrichment.*

Set	PRS.R2	Coefficient (Direction)	P	Num_SNP	Competitive.P
REACTOME EPH EPHRIN SIGNALING	0.0097	674	0.00003	822	0.0004
WP INTEGRIN MEDIATED CELL ADHESION	0.0059	526	0.00111	804	0.0051
REACTOME EPHA MEDIATED GROWTH CONE COLLAPSE	0.0052	272	0.00238	258	0.0062
REACTOME FCGAMMA RECEPTOR FCGR DEPENDENT PHAGOCYTOSIS	0.0057	508	0.00139	846	0.0066
REACTOME INFECTIOUS DISEASE	0.0095	1592	0.00003	4504	0.0066
KEGG MEDICUS REFERENCE REGULATION OF GF RTK RAS ERK SIGNALING RAS UBIQUITINATION BY CUL3 COMPLEX	0.0041	54	0.00690	21	0.0073
Immune response to viral infections	0.0095	1592	0.00003	4504	0.0084
WP SPHINGOLIPID METABOLISM INTEGRATED PATHWAY	0.0041	221	0.00665	197	0.0107
WP CONGENITAL GENERALIZED LIPODYSTROPHY	0.0038	123	0.00933	80	0.0110
Protein Localization and Fatty acid metabolism	0.0056	642	0.00149	1057	0.0115
WP EICOSANOID METABOLISM VIA LIPOOXYGENASES LOX	0.0039	156	0.00853	130	0.0116
REACTOME PARASITE INFECTION	0.0046	346	0.00398	513	0.0120
REACTOME ELASTIC FIBRE FORMATION	0.0041	260	0.00658	356	0.0138
PID EPHA2 FWD PATHWAY	0.0037	195	0.00959	172	0.0150
REACTOME RHO GTPASE CYCLE	0.0077	1211	0.00021	3501	0.0150
REACTOME SIGNALING BY RHO GTPASES MIRO GTPASES AND RHOBTB3	0.0081	1446	0.00014	4440	0.0169
REACTOME MHC CLASS II ANTIGEN PRESENTATION	0.0041	386	0.00663	489	0.0175
KEGG MEDICUS REFERENCE MICROTUBULE DEPOLYMERIZATION	0.0034	130	0.01416	66	0.0180
REACTOME HIV INFECTION	0.0048	536	0.00337	991	0.0197
REACTOME MOLECULES ASSOCIATED WITH ELASTIC FIBRES	0.0036	222	0.01171	315	0.0209
REACTOME PROTEIN LOCALIZATION	0.0041	437	0.00696	639	0.0221
REACTOME PRE NOTCH PROCESSING IN GOLGI	0.0031	108	0.01844	79	0.0241
PID AVB3 OPN PATHWAY	0.0032	210	0.01611	255	0.0250
WP SPHINGOLIPID METABOLISM OVERVIEW	0.0033	197	0.01582	201	0.0251
REACTOME CHYLOMICRON CLEARANCE	0.0028	78	0.02506	39	0.0296
KEGG PEROXISOME	0.0032	282	0.01662	376	0.0314
KEGG MEDICUS REFERENCE LPA GNA12 13 RHOA SIGNALING PATHWAY	0.0028	105	0.02567	75	0.0323
Metabolism of nucleotide sugars and glycolysis	0.0035	320	0.01254	484	0.0327
REACTOME VIRAL INFECTION PATHWAYS	0.0061	1098	0.00098	3235	0.0330

REACTOME EPIGENETIC REGULATION OF GENE EXPRESSION	0.0034	366	0.01382	552	0.0340
KEGG AMINO SUGAR AND NUCLEOTIDE SUGAR METABOLISM	0.0029	209	0.02267	257	0.0346
REACTOME SIGNALING BY MAPK MUTANTS	0.0026	-76	0.03076	31	0.0346
WP PARKIN UBIQUITIN PROTEASOMAL SYSTEM PATHWAY	0.0029	206	0.02284	207	0.0360
WP DISORDERS OF GALACTOSE METABOLISM	0.0026	91	0.03158	69	0.0409
PID EPHA FWDPATHWAY	0.0029	229	0.02357	326	0.0411
WP GENES RELATED TO PRIMARY CILIUM DEVELOPMENT BASED ON CRISPR	0.0031	292	0.01943	512	0.0432
REACTOME GAP JUNCTION TRAFFICKING AND REGULATION	0.0026	188	0.03041	171	0.0445
REACTOME SIGNALING BY VEGF	0.0034	422	0.01361	873	0.0465
Sphingolipid metabolism	0.0025	218	0.03291	308	0.0553
WP SPHINGOLIPID PATHWAY	0.0023	187	0.04281	238	0.0660
REACTOME GABA SYNTHESIS RELEASE REUPTAKE AND DEGRADATION	0.0020	138	0.05574	158	0.0712
KEGG MEDICUS REFERENCE CA2 PLCD ITPR SIGNALING PATHWAY	0.0020	94	0.05899	110	0.0715
REACTOME EARLY PHASE OF HIV LIFE CYCLE	0.0018	89	0.07245	58	0.0830
REACTOME COENZYME A BIOSYNTHESIS	0.0017	36	0.08537	33	0.0917
REACTOME MITOCHONDRIAL TRANSLATION	0.0019	191	0.06505	302	0.0994
REACTOME PINK1 PRKN MEDIATED MITOPHAGY	0.0017	85	0.08397	73	0.0997
REACTOME CDC42 GTPASE CYCLE	0.0030	478	0.02107	1556	0.1020
BIOCARTA SAM68 PATHWAY	0.0015	-48	0.09733	22	0.1046
KEGG MEDICUS REFERENCE REGULATION OF GF RTK RAS ERK SIGNALING RASGAP	0.0014	50	0.10755	32	0.1141
Extracellular Matrix	0.0041	838	0.00696	3375	0.1145
REACTOME PEROXISOMAL PROTEIN IMPORT	0.0016	192	0.08838	306	0.1277
KEGG SPHINGOLIPID METABOLISM	0.0015	145	0.09747	222	0.1293
REACTOME CILIUM ASSEMBLY	0.0021	339	0.05398	917	0.1313
Organelle biosynthesis	0.0027	581	0.02659	1986	0.1425
WP MIRNA TARGETS IN ECM AND MEMBRANE RECEPTORS	0.0015	142	0.10725	257	0.1450
REACTOME REGULATION OF HMOX1 EXPRESSION AND ACTIVITY	0.0012	-45	0.13953	32	0.1465
KEGG GALACTOSE METABOLISM	0.0013	104	0.12813	149	0.1475
WP EICOSANOID METABOLISM VIA CYCLOOXYGENASES COX	0.0013	123	0.12599	197	0.1546
WP VEGFA VEGFR2 SIGNALING	0.0031	730	0.01771	2790	0.1602
REACTOME SIGNALING BY PDGF	0.0014	190	0.10973	410	0.1622
Signal Transduction	0.0040	1054	0.00729	4477	0.1672
REACTOME HIV LIFE CYCLE	0.0015	247	0.10365	593	0.1675

REACTOME RUNX2 REGULATES OSTEOBLAST DIFFERENTIATION	0.0012	-101	0.14230	172	0.1714
REACTOME EXTRACELLULAR MATRIX ORGANIZATION	0.0028	589	0.02487	2468	0.1725
REACTOME DISEASES ASSOCIATED WITH GLYCOSYLATION PRECURSOR BIOSYNTHESIS	0.0011	55	0.15344	62	0.1729
REACTOME ANCHORING OF THE BASAL BODY TO THE PLASMA MEMBRANE	0.0013	188	0.12200	454	0.1783
REACTOME TCF DEPENDENT SIGNALING IN RESPONSE TO WNT	0.0016	305	0.09368	924	0.1902
REACTOME CALCITONIN LIKE LIGAND RECEPTORS	0.0010	-46	0.17999	46	0.1948
REACTOME METABOLISM OF CARBOHYDRATES	0.0021	503	0.05050	1799	0.2000
KEGG UBIQUITIN MEDIATED PROTEOLYSIS	0.0012	220	0.13916	562	0.2100
REACTOME PEROXISOMAL LIPID METABOLISM	0.0010	97	0.18351	169	0.2126
WP HEDGEHOG SIGNALING PATHWAY WP4249	0.0010	132	0.17135	294	0.2169
WP ESTROGEN RECEPTOR PATHWAY	0.0009	58	0.19328	85	0.2181
Carbohydrate metabolism	0.0021	506	0.05509	1875	0.2209
REACTOME MITOPHAGY	0.0009	74	0.19850	94	0.2267
BIOCARTA PARKIN PATHWAY	0.0008	28	0.22445	29	0.2282
REACTOME ORGANELLE BIOGENESIS AND MAINTENANCE	0.0015	350	0.09894	1297	0.2416
REACTOME SIGNALING BY RECEPTOR TYROSINE KINASES	0.0029	872	0.02221	4214	0.2675
Inflammatory signaling	0.0009	154	0.19848	472	0.2695
WP BARDET BIEDL SYNDROME	0.0009	170	0.20655	561	0.2860
WP RESISTIN AS A REGULATOR OF INFLAMMATION	0.0008	120	0.24688	341	0.3017
BIOCARTA HIF PATHWAY	0.0006	48	0.28251	63	0.3067
BIOCARTA ACTINY PATHWAY	0.0006	40	0.28948	52	0.3128
REACTOME DEFECTIVE CHSY1 CAUSES TPBS	0.0006	29	0.29709	42	0.3146
KEGG MEDICUS REFERENCE PROK PRKR GI ERK SIGNALING PATHWAY	0.0006	42	0.30036	43	0.3203
PID TOLL ENDOGENOUS PATHWAY	0.0006	58	0.29448	100	0.3289
WP ONE CARBON METABOLISM	0.0006	-85	0.30614	197	0.3383
REACTOME DEFECTIVE CHST14 CAUSES EDS MUSCULOCONTRACTURAL TYPE	0.0005	23	0.33921	35	0.3490
KEGG MEDICUS VARIANT MUTATION INACTIVATED PROK2 TO PROK PRKR GI ERK SIGNALING PATHWAY	0.0005	36	0.33630	39	0.3505
REACTOME DEFECTIVE B4GALT7 CAUSES EDS PROGEROID TYPE	0.0006	-83	0.31183	227	0.3529
REACTOME CS DS DEGRADATION	0.0005	36	0.35156	72	0.3816
REACTOME NEGATIVE REGULATION OF MAPK PATHWAY	0.0005	-76	0.36304	194	0.3956

WP AEROBIC GLYCOLYSIS	0.0004	41	0.37485	53	0.4000
WP FAMILIAL HYPERLIPIDEMIA TYPE 3	0.0004	49	0.37922	87	0.4096
WP FOCAL ADHESION	0.0009	294	0.20895	1694	0.4305
REACTOME VITAMIN B5 PANTOTHENATE METABOLISM	0.0004	34	0.40248	89	0.4324
REACTOME MRNA DECAY BY 5 TO 3 EXORIBONUCLEASE	0.0004	42	0.42508	58	0.4454
REACTOME SPERM MOTILITY AND TAXES	0.0003	29	0.45612	60	0.4791
REACTOME TRANSCRIPTION OF E2F TARGETS UNDER NEGATIVE CONTROL BY DREAM COMPLEX	0.0003	38	0.47703	67	0.5015
REACTOME DEFECTIVE CHST3 CAUSES SEDCJD	0.0003	18	0.49604	40	0.5061
REACTOME DERMATAN SULFATE BIOSYNTHESIS	0.0003	37	0.49640	101	0.5256
REACTOME NEGATIVE REGULATION OF TCF DEPENDENT SIGNALING BY DVL INTERACTING PROTEINS	0.0002	15	0.51927	34	0.5268
REACTOME VITAMINS	0.0002	-17	0.53797	19	0.5444
BIOCARTA NPP1 PATHWAY	0.0002	33	0.51225	115	0.5463
REACTOME TRIGLYCERIDE BIOSYNTHESIS	0.0002	32	0.54965	77	0.5784
KEGG MEDICUS REFERENCE NON HOMOLOGOUS END JOINING	0.0002	-29	0.57227	64	0.5916
BIOCARTA PRION PATHWAY	0.0002	27	0.58956	95	0.6125
WP MALE STEROID HORMONES IN CARDIOMYOCYTE ENERGY METABOLISM	0.0002	25	0.60167	67	0.6229
REACTOME LAMININ INTERACTIONS	0.0002	50	0.60209	309	0.6383
Non homologous end joining	0.0001	34	0.61966	112	0.6410
KEGG MEDICUS VARIANT MUTATION INACTIVATED P62 TO PINK PARKIN MEDIATED AUTOPHAGOSOME FORMATION	0.0001	17	0.67080	57	0.6834
REACTOME ALPHA LINOLENIC OMEGA3 AND LINOLEIC OMEGA6 ACID METABOLISM	0.0001	16	0.66558	67	0.6881
REACTOME G0 AND EARLY G1	0.0001	-27	0.68224	105	0.7022
WP GLYCOLYSIS AND GLUCONEOGENESIS	0.0001	34	0.69006	171	0.7104
PID BARD1 PATHWAY	0.0001	29	0.69396	127	0.7159
KEGG MEDICUS REFERENCE V D J RECOMBINATION	0.0001	-18	0.70841	57	0.7190
KEGG ALPHA LINOLENIC ACID METABOLISM	0.0001	-17	0.71182	89	0.7325
WP HEDGEHOG SIGNALING PATHWAY WP47	0.0001	-15	0.73560	81	0.7506
REACTOME ASSEMBLY OF COLLAGEN FIBRILS AND OTHER MULTIMERIC STRUCTURES	0.0001	50	0.70867	513	0.7564
REACTOME RESPIRATORY ELECTRON TRANSPORT	0.0001	39	0.73422	362	0.7647
KEGG MEDICUS REFERENCE PINK PARKIN MEDIATED AUTOPHAGOSOME FORMATION	0.0000	12	0.79969	79	0.8143

REACTOME RESPONSE OF MTB TO PHAGOCYTOSIS	0.0000	-14	0.80616	86	0.8167
REACTOME PLASMA LIPOPROTEIN REMODELING	0.0000	23	0.81189	267	0.8237
KEGG MEDICUS REFERENCE ACTIVATION OF PRC2.2 BY UBIQUITINATION OF H2AK119 IN GERMLINE GENES	0.0000	-16	0.82477	132	0.8349
KEGG MEDICUS VARIANT MUTATION INACTIVATED OPTN TO PINK PARKIN MEDIATED AUTOPHAGOSOME FORMATION	0.0000	-6	0.84967	43	0.8571
PID AVB3 INTEGRIN PATHWAY	0.0000	27	0.84643	606	0.8723
REACTOME A TETRASACCHARIDE LINKER SEQUENCE IS REQUIRED FOR GAG SYNTHESIS	0.0000	-14	0.88244	290	0.8923
KEGG NON HOMOLOGOUS END JOINING	0.0000	-5	0.90812	56	0.9097
REACTOME DISEASES ASSOCIATED WITH GLYCOSAMINOGLYCAN METABOLISM	0.0000	-13	0.89918	341	0.9120
WP BIOSYNTHESIS AND TURNOVER OF 1 DEOXY SPHINGOID BASES	0.0000	-6	0.90801	64	0.9137
WP 2Q11 2 COPY NUMBER VARIATION SYNDROME	0.0000	4	0.93485	54	0.9342
KEGG MEDICUS ENV FACTOR IRON TO ANTEROGRADE AXONAL TRANSPORT	0.0000	-3	0.93579	42	0.9386
REACTOME TRANSCRIPTIONAL REGULATION BY E2F6	0.0000	4	0.95025	133	0.9528
WP OMEGA 3 OMEGA 6 FATTY ACID SYNTHESIS	0.0000	3	0.95609	88	0.9588
REACTOME PEPTIDE HORMONE BIOSYNTHESIS	0.0000	1	0.97240	29	0.9738
REACTOME MITOCHONDRIAL RNA DEGRADATION	0.0000	1	0.97751	53	0.9766
KEGG SMALL CELL LUNG CANCER	0.0000	-3	0.98696	753	0.9904

Supplementary Table 7. *Meta-analysis of WMH ps-PRS association with WMH volume across UK Biobank, 3C-Dijon, Memento, and i-Share for WMH ps-PRS enriched in at least 1 follow-up cohort. Green indicates passing multiple testing correction.*

PRS	**BETA_Meta**	**P_Meta**	**Q_pval**
Signal Transduction	0.065 [0.052, 0.077]	6.93E-24	0.5932
Organelle biosynthesis	0.06 [0.048, 0.072]	1.46E-21	0.4804
REACTOME ANCHORING OF THE BASAL BODY TO THE PLASMA MEMBRANE	0.047 [0.035, 0.058]	3.39E-15	0.6449
REACTOME ORGANELLE BIOGENESIS AND MAINTENANCE	0.048 [0.035, 0.06]	4.95E-14	0.7141
REACTOME SIGNALING BY RECEPTOR TYROSINE KINASES	0.059 [0.043, 0.075]	1.06E-13	0.5222
WP FOCAL ADHESION	0.043 [0.032, 0.055]	3.46E-13	0.5841
REACTOME TCF DEPENDENT SIGNALING IN RESPONSE TO WNT	0.041 [0.029, 0.052]	7.30E-12	0.7313
WP GENES RELATED TO PRIMARY CILIUM DEVELOPMENT BASED ON CRISPR	0.043 [0.031, 0.055]	7.39E-12	0.6284
WP INTEGRIN MEDIATED CELL ADHESION	0.041 [0.029, 0.053]	8.76E-12	0.4115
Protein Localization and Fatty acid metabolism	0.042 [0.03, 0.055]	6.10E-11	0.2704
REACTOME PROTEIN LOCALIZATION	0.038 [0.027, 0.05]	9.35E-11	0.6172
REACTOME SIGNALING BY PDGF	0.039 [0.027, 0.051]	6.91E-10	0.6143
WP CONGENITAL GENERALIZED LIPODYSTROPHY	0.037 [0.025, 0.048]	7.21E-10	0.6790
WP SPHINGOLIPID METABOLISM OVERVIEW	0.036 [0.024, 0.048]	1.54E-09	0.4400
REACTOME VITAMIN B5 PANTOTHENATE METABOLISM	0.037 [0.025, 0.049]	4.96E-09	0.5636
REACTOME EPIGENETIC REGULATION OF GENE EXPRESSION	0.037 [0.025, 0.049]	5.20E-09	0.3249
REACTOME MITOCHONDRIAL TRANSLATION	0.035 [0.022, 0.047]	5.73E-08	0.1577
WP RESISTIN AS A REGULATOR OF INFLAMMATION	0.032 [0.021, 0.044]	5.87E-08	0.6886
KEGG PEROXISOME	0.032 [0.02, 0.043]	1.14E-07	0.6113
REACTOME MHC CLASS II ANTIGEN PRESENTATION	0.034 [0.021, 0.046]	1.20E-07	0.3392
KEGG MEDICUS REFERENCE REGULATION OF GF RTK RAS ERK SIGNALING RAS UBIQUITINATION BY CUL3 COMPLEX	0.033 [0.021, 0.046]	1.25E-07	0.7092
KEGG MEDICUS REFERENCE MICROTUBULE DEPOLYMERIZATION	0.033 [0.021, 0.046]	1.62E-07	0.4033
REACTOME MOLECULES ASSOCIATED WITH ELASTIC FIBRES	0.045 [0.028, 0.061]	2.23E-07	0.1842
REACTOME GAP JUNCTION TRAFFICKING AND REGULATION	0.032 [0.02, 0.045]	3.65E-07	0.4299
PID EPHA FWDPATHWAY	0.032 [0.019, 0.044]	4.91E-07	0.7497
WP OMEGA 3 OMEGA 6 FATTY ACID SYNTHESIS	0.031 [0.019, 0.044]	1.03E-06	0.3816
WP DISORDERS OF GALACTOSE METABOLISM	0.03 [0.018, 0.043]	1.44E-06	0.8862
WP EICOSANOID METABOLISM VIA LIPOOXYGENASES LOX	0.029 [0.017, 0.04]	1.59E-06	0.6623

REACTOME CHYLOMICRON CLEARANCE	0.029 [0.017, 0.041]	3.06E-06	0.2917
Metabolism of nucleotide sugars and glycolysis	0.041 [0.024, 0.059]	4.15E-06	0.2910
BIOCARTA NPP1 PATHWAY	0.028 [0.016, 0.04]	9.73E-06	0.6144
REACTOME VIRAL INFECTION PATHWAYS	0.051 [0.028, 0.074]	1.17E-05	0.1448
PID EPHA2 FWD PATHWAY	0.034 [0.017, 0.05]	6.27E-05	0.4371
REACTOME HIV INFECTION	0.041 [0.02, 0.063]	1.44E-04	0.1872
WP SPHINGOLIPID METABOLISM INTEGRATED PATHWAY	0.032 [0.015, 0.05]	2.10E-04	0.2111
REACTOME ALPHA LINOLENIC OMEGA3 AND LINOLEIC OMEGA6 ACID METABOLISM	0.039 [0.015, 0.062]	1.34E-03	0.1374
REACTOME ELASTIC FIBRE FORMATION	0.041 [0.015, 0.067]	1.68E-03	0.1032
Immune response to viral infections	0.054 [0.02, 0.088]	1.77E-03	0.0284
REACTOME INFECTIOUS DISEASE	0.054 [0.02, 0.088]	1.77E-03	0.0284
WP ESTROGEN RECEPTOR PATHWAY	0.048 [0.018, 0.078]	1.90E-03	0.0527
KEGG SMALL CELL LUNG CANCER	0.03 [0.011, 0.05]	2.32E-03	0.2340
REACTOME EPHA MEDIATED GROWTH CONE COLLAPSE	0.028 [0.009, 0.047]	3.09E-03	0.1892
REACTOME EPH EPHRIN SIGNALING	0.041 [0.014, 0.069]	3.11E-03	0.0902
REACTOME SIGNALING BY RHO GTPASES MIRO GTPASES AND RHOBTB3	0.045 [0.014, 0.076]	4.70E-03	0.0465
REACTOME RHO GTPASE CYCLE	0.04 [0.007, 0.074]	1.64E-02	0.0322
PID AVB3 OPN PATHWAY	0.025 [0.004, 0.047]	1.84E-02	0.1757
REACTOME TRANSCRIPTION OF E2F TARGETS UNDER NEGATIVE CONTROL BY DREAM COMPLEX	0.032 [0.005, 0.059]	1.89E-02	0.0916
REACTOME DISEASES ASSOCIATED WITH GLYCOSAMINOGLYCAN METABOLISM	0.026 [0.003, 0.049]	2.86E-02	0.1429
Carbohydrate metabolism	0.033 [0.002, 0.064]	3.50E-02	0.0420
REACTOME SIGNALING BY VEGF	0.028 [0, 0.056]	4.65E-02	0.0730
KEGG AMINO SUGAR AND NUCLEOTIDE SUGAR METABOLISM	0.025 [-0.002, 0.051]	6.47E-02	0.0998
REACTOME G0 AND EARLY G1	0.03 [-0.006, 0.066]	9.73E-02	0.0260
REACTOME PRE NOTCH PROCESSING IN GOLGI	0.019 [-0.004, 0.042]	1.11E-01	0.1463
REACTOME FCGAMMA RECEPTOR FCGR DEPENDENT PHAGOCYTOSIS	0.028 [-0.007, 0.063]	1.18E-01	0.0277
KEGG MEDICUS REFERENCE ACTIVATION OF PRC2.2 BY UBIQUITINATION OF H2AK119 IN GERMLINE GENES	0.019 [-0.006, 0.044]	1.35E-01	0.1048
WP HEDGEHOG SIGNALING PATHWAY WP4249	0.021 [-0.011, 0.053]	1.97E-01	0.0488
REACTOME CALCITONIN LIKE LIGAND RECEPTORS	0.019 [-0.01, 0.049]	2.02E-01	0.0602
REACTOME PARASITE INFECTION	0.02 [-0.016, 0.055]	2.77E-01	0.0272
KEGG MEDICUS REFERENCE LPA GNA12 13 RHOA SIGNALING PATHWAY	0.017 [-0.017, 0.05]	3.35E-01	0.0258
WP PARKIN UBIQUITIN PROTEASOMAL SYSTEM PATHWAY	0.014 [-0.018, 0.046]	3.91E-01	0.0411
PID TOLL ENDOGENOUS PATHWAY	0.014 [-0.027, 0.056]	4.95E-01	0.0079

Supplementary Table 8. *Associations of WMH ps-PRS with WMH volume for those enriched in follow-up analysis.*

PRS	BETA_UKB	P_UKB	BETA_3CDijon	P_3CDijon	BETA_MEMENTO	P_MEMENTO	BETA_iSHARE	P_iSHARE
Signal Transduction	0.069 [0.055, 0.083]	1.75E-23	0.038 [-0.012, 0.087]	1.34E-01	0.055 [0.015, 0.096]	7.29E-03	0.054 [0.005, 0.103]	3.22E-02
Organelle biosynthesis	0.063 [0.05, 0.077]	7.11E-20	0.076 [0.026, 0.125]	2.66E-03	0.045 [0.005, 0.085]	2.66E-02	0.031 [-0.019, 0.08]	2.29E-01
REACTOME ANCHORING OF THE BASAL BODY TO THE PLASMA MEMBRANE	0.046 [0.032, 0.059]	3.01E-11	0.055 [0.005, 0.104]	2.97E-02	0.032 [-0.008, 0.072]	1.22E-01	0.072 [0.022, 0.121]	4.68E-03
REACTOME ORGANELLE BIOGENESIS AND MAINTENANCE	0.048 [0.034, 0.062]	3.67E-12	0.07 [0.02, 0.119]	5.62E-03	0.034 [-0.006, 0.074]	9.89E-02	0.039 [-0.01, 0.089]	1.22E-01
REACTOME SIGNALING BY RECEPTOR TYROSINE KINASES	0.066 [0.053, 0.08]	1.01E-21	0.034 [-0.015, 0.084]	1.72E-01	0.047 [0.007, 0.088]	2.22E-02	0.05 [0.001, 0.099]	4.78E-02
WP FOCAL ADHESION	0.045 [0.031, 0.058]	9.29E-11	0.025 [-0.025, 0.074]	3.30E-01	0.026 [-0.015, 0.067]	2.09E-01	0.064 [0.015, 0.114]	1.13E-02
REACTOME TCF DEPENDENT SIGNALING IN RESPONSE TO WNT	0.04 [0.026, 0.053]	9.28E-09	0.033 [-0.016, 0.082]	1.87E-01	0.035 [-0.006, 0.075]	9.37E-02	0.067 [0.017, 0.116]	8.06E-03
WP GENES RELATED TO PRIMARY CILIUM DEVELOPMENT BASED ON CRISPR	0.046 [0.033, 0.06]	1.86E-11	0.023 [-0.026, 0.073]	3.55E-01	0.048 [0.008, 0.087]	1.94E-02	0.019 [-0.03, 0.069]	4.41E-01
WP INTEGRIN MEDIATED CELL ADHESION	0.041 [0.027, 0.054]	3.40E-09	0.017 [-0.032, 0.067]	4.90E-01	0.067 [0.027, 0.108]	1.11E-03	0.024 [-0.025, 0.074]	3.35E-01
Protein Localization and Fatty acid metabolism	0.041 [0.028, 0.055]	2.62E-09	0.005 [-0.044, 0.055]	8.33E-01	0.065 [0.025, 0.104]	1.49E-03	0.059 [0.009, 0.108]	2.00E-02
REACTOME PROTEIN LOCALIZATION	0.038 [0.024, 0.051]	4.31E-08	0.013 [-0.036, 0.062]	6.11E-01	0.055 [0.015, 0.095]	6.96E-03	0.046 [-0.003, 0.095]	6.65E-02
REACTOME SIGNALING BY PDGF	0.037 [0.023, 0.051]	8.41E-08	0.071 [0.022, 0.12]	4.88E-03	0.033 [-0.007, 0.073]	1.10E-01	0.039 [-0.01, 0.089]	1.19E-01
WP CONGENITAL GENERALIZED LIPODYSTROPHY	0.037 [0.023, 0.05]	1.01E-07	0.028 [-0.022, 0.077]	2.69E-01	0.053 [0.013, 0.093]	9.33E-03	0.015 [-0.035, 0.064]	5.60E-01
WP SPHINGOLIPID METABOLISM OVERVIEW	0.038 [0.024, 0.051]	4.64E-08	0.003 [-0.046, 0.052]	8.96E-01	0.049 [0.009, 0.089]	1.58E-02	0.019 [-0.03, 0.068]	4.52E-01
REACTOME VITAMIN B5 PANTOTHENATE METABOLISM	0.038 [0.025, 0.052]	3.28E-08	0.029 [-0.02, 0.078]	2.52E-01	0.017 [-0.023, 0.057]	4.02E-01	0.062 [0.013, 0.112]	1.36E-02
REACTOME EPIGENETIC REGULATION OF GENE EXPRESSION	0.04 [0.027, 0.054]	5.77E-09	0.001 [-0.049, 0.05]	9.81E-01	0.05 [0.01, 0.09]	1.38E-02	0.014 [-0.035, 0.064]	5.71E-01

REACTOME MITOCHONDRIAL TRANSLATION	0.034 [0.021, 0.048]	6.44E-07	0.077 [0.028, 0.126]	2.11E-03	0.038 [-0.002, 0.078]	6.51E-02	-0.004 [-0.053, 0.046]	8.82E-01
WP RESISTIN AS A REGULATOR OF INFLAMMATION	0.033 [0.019, 0.046]	2.44E-06	0.013 [-0.036, 0.063]	5.92E-01	0.024 [-0.016, 0.064]	2.47E-01	0.054 [0.004, 0.104]	3.30E-02
KEGG PEROXISOME	0.03 [0.016, 0.043]	1.85E-05	0.011 [-0.038, 0.061]	6.50E-01	0.049 [0.009, 0.089]	1.66E-02	0.047 [-0.003, 0.096]	6.50E-02
REACTOME MHC CLASS II ANTIGEN PRESENTATION	0.035 [0.021, 0.048]	5.11E-07	0.003 [-0.047, 0.052]	9.21E-01	0.056 [0.015, 0.096]	6.63E-03	0.014 [-0.035, 0.063]	5.76E-01
KEGG MEDICUS REFERENCE REGULATION OF GF RTK RAS ERK SIGNALING RAS UBIQUITINATION BY CUL3 COMPLEX	0.032 [0.019, 0.046]	2.91E-06	0.025 [-0.024, 0.074]	3.21E-01	0.055 [0.015, 0.095]	6.90E-03	0.025 [-0.025, 0.074]	3.27E-01
KEGG MEDICUS REFERENCE MICROTUBULE DEPOLYMERIZATION	0.032 [0.018, 0.046]	3.78E-06	0.055 [0.005, 0.104]	2.99E-02	0.05 [0.01, 0.09]	1.42E-02	0.003 [-0.047, 0.052]	9.09E-01
REACTOME MOLECULES ASSOCIATED WITH ELASTIC FIBRES	0.049 [0.035, 0.062]	2.17E-12	0.058 [0.009, 0.107]	2.15E-02	0.052 [0.012, 0.092]	1.17E-02	-0.006 [-0.056, 0.043]	8.04E-01
REACTOME GAP JUNCTION TRAFFICKING AND REGULATION	0.034 [0.021, 0.048]	7.52E-07	0.031 [-0.019, 0.08]	2.24E-01	0.044 [0.004, 0.084]	3.04E-02	-0.007 [-0.056, 0.043]	7.96E-01
PID EPHA FWDPATHWAY	0.032 [0.019, 0.046]	3.51E-06	0.027 [-0.022, 0.076]	2.86E-01	0.046 [0.006, 0.086]	2.36E-02	0.011 [-0.039, 0.06]	6.72E-01
WP OMEGA 3 OMEGA 6 FATTY ACID SYNTHESIS	0.033 [0.02, 0.047]	1.31E-06	0.037 [-0.012, 0.086]	1.39E-01	0.001 [-0.039, 0.041]	9.56E-01	0.053 [0.004, 0.102]	3.46E-02
WP DISORDERS OF GALACTOSE METABOLISM	0.03 [0.017, 0.044]	1.20E-05	0.023 [-0.026, 0.073]	3.50E-01	0.044 [0.004, 0.084]	3.16E-02	0.022 [-0.028, 0.071]	3.86E-01
WP EICOSANOID METABOLISM VIA LIPOOXYGENASES LOX	0.026 [0.012, 0.039]	1.98E-04	0.029 [-0.02, 0.078]	2.44E-01	0.053 [0.014, 0.093]	8.54E-03	0.027 [-0.022, 0.076]	2.81E-01
REACTOME CHYLOMICRON CLEARANCE	0.031 [0.017, 0.044]	8.19E-06	0.009 [-0.04, 0.058]	7.16E-01	0.046 [0.006, 0.086]	2.51E-02	-0.009 [-0.059, 0.04]	7.16E-01
Metabolism of nucleotide sugars and glycolysis	0.049 [0.035, 0.062]	1.79E-12	0.015 [-0.035, 0.064]	5.60E-01	0.051 [0.011, 0.091]	1.25E-02	0.011 [-0.039, 0.06]	6.73E-01
BIOCARTA NPP1 PATHWAY	0.028 [0.015, 0.042]	4.70E-05	0.022 [-0.027, 0.071]	3.87E-01	0.013 [-0.027, 0.053]	5.12E-01	0.056 [0.007, 0.106]	2.57E-02
REACTOME VIRAL INFECTION PATHWAYS	0.06 [0.047, 0.074]	2.84E-18	0.049 [-0.001, 0.098]	5.37E-02	0.068 [0.027, 0.108]	9.84E-04	0.002 [-0.048, 0.051]	9.42E-01
PID EPHA2 FWD PATHWAY	0.026 [0.012, 0.039]	2.04E-04	0.032 [-0.017, 0.082]	2.04E-01	0.053 [0.013, 0.093]	9.59E-03	0.056 [0.007, 0.105]	2.67E-02

REACTOME HIV INFECTION	0.051 [0.038, 0.065]	1.44E-13	0.014 [-0.036, 0.063]	5.91E-01	0.06 [0.02, 0.1]	3.37E-03	0.008 [-0.041, 0.058]	7.43E-01
WP SPHINGOLIPID METABOLISM INTEGRATED PATHWAY	0.037 [0.023, 0.05]	1.03E-07	-0.006 [-0.055, 0.043]	8.14E-01	0.055 [0.015, 0.095]	6.65E-03	0.013 [-0.037, 0.062]	6.16E-01
REACTOME ALPHA LINOLENIC OMEGA3 AND LINOLEIC OMEGA6 ACID METABOLISM	0.032 [0.019, 0.046]	2.62E-06	0.078 [0.028, 0.127]	2.07E-03	0.009 [-0.031, 0.049]	6.66E-01	0.057 [0.007, 0.106]	2.43E-02
REACTOME ELASTIC FIBRE FORMATION	0.047 [0.034, 0.061]	8.79E-12	0.062 [0.013, 0.111]	1.38E-02	0.056 [0.016, 0.096]	6.59E-03	-0.014 [-0.063, 0.036]	5.81E-01
Immune response to viral infections	0.071 [0.058, 0.085]	6.47E-25	0.044 [-0.005, 0.094]	7.71E-02	0.085 [0.045, 0.125]	3.47E-05	0 [-0.05, 0.049]	9.87E-01
REACTOME INFECTIOUS DISEASE	0.071 [0.058, 0.085]	6.47E-25	0.044 [-0.005, 0.094]	7.71E-02	0.085 [0.045, 0.125]	3.47E-05	0 [-0.05, 0.049]	9.87E-01
WP ESTROGEN RECEPTOR PATHWAY	0.033 [0.019, 0.046]	1.89E-06	0.103 [0.054, 0.152]	3.76E-05	0.027 [-0.013, 0.067]	1.93E-01	0.046 [-0.003, 0.096]	6.57E-02
KEGG SMALL CELL LUNG CANCER	0.038 [0.024, 0.051]	4.54E-08	0.016 [-0.033, 0.065]	5.26E-01	0 [-0.04, 0.04]	9.87E-01	0.054 [0.005, 0.104]	3.04E-02
REACTOME EPHA MEDIATED GROWTH CONE COLLAPSE	0.028 [0.015, 0.042]	3.82E-05	-0.004 [-0.053, 0.045]	8.81E-01	0.062 [0.022, 0.102]	2.38E-03	0.013 [-0.036, 0.062]	6.07E-01
REACTOME EPH EPHRIN SIGNALING	0.037 [0.023, 0.051]	8.94E-08	0.022 [-0.027, 0.072]	3.79E-01	0.085 [0.045, 0.125]	3.12E-05	0.015 [-0.034, 0.065]	5.39E-01
REACTOME SIGNALING BY RHO GTPASES MIRO GTPASES AND RHOBTB3	0.06 [0.046, 0.073]	5.63E-18	0.008 [-0.042, 0.057]	7.65E-01	0.078 [0.038, 0.119]	1.40E-04	0.014 [-0.036, 0.063]	5.85E-01
REACTOME RHO GTPASE CYCLE	0.057 [0.044, 0.071]	1.16E-16	0.005 [-0.045, 0.054]	8.55E-01	0.076 [0.036, 0.116]	2.11E-04	0.006 [-0.044, 0.055]	8.19E-01
PID AVB3 OPN PATHWAY	0.033 [0.019, 0.046]	2.05E-06	-0.013 [-0.063, 0.036]	5.93E-01	0.049 [0.009, 0.09]	1.61E-02	0.005 [-0.044, 0.055]	8.31E-01
REACTOME TRANSCRIPTION OF E2F TARGETS UNDER NEGATIVE CONTROL BY DREAM COMPLEX	0.029 [0.016, 0.043]	2.16E-05	0.086 [0.037, 0.135]	6.21E-04	0.015 [-0.025, 0.054]	4.77E-01	0.007 [-0.043, 0.056]	7.90E-01
REACTOME DISEASES ASSOCIATED WITH GLYCOSAMINOGLYCAN METABOLISM	0.035 [0.021, 0.048]	5.62E-07	0.002 [-0.047, 0.051]	9.38E-01	-0.003 [-0.043, 0.038]	8.99E-01	0.057 [0.007, 0.106]	2.42E-02
Carbohydrate metabolism	0.051 [0.038, 0.065]	1.27E-13	-0.023 [-0.072, 0.026]	3.57E-01	0.039 [-0.001, 0.079]	5.51E-02	0.048 [-0.001, 0.098]	5.39E-02
REACTOME SIGNALING BY VEGF	0.045 [0.032, 0.059]	7.03E-11	-0.004 [-0.053, 0.045]	8.79E-01	0.051 [0.01, 0.091]	1.36E-02	-0.004 [-0.054, 0.045]	8.61E-01

KEGG AMINO SUGAR AND NUCLEOTIDE SUGAR METABOLISM	0.039 [0.026, 0.053]	1.21E-08	-0.014 [-0.063, 0.035]	5.77E-01	0.047 [0.007, 0.087]	2.27E-02	0.001 [-0.048, 0.051]	9.53E-01
REACTOME G0 AND EARLY G1	0.033 [0.019, 0.047]	1.84E-06	0.087 [0.038, 0.136]	5.05E-04	-0.008 [-0.048, 0.032]	6.82E-01	0.013 [-0.037, 0.062]	6.12E-01
REACTOME PRE NOTCH PROCESSING IN GOLGI	0.026 [0.013, 0.04]	1.41E-04	-0.01 [-0.059, 0.039]	6.87E-01	0.048 [0.008, 0.088]	1.84E-02	-0.012 [-0.061, 0.037]	6.30E-01
REACTOME FCGAMMA RECEPTOR FCGR DEPENDENT PHAGOCYTOSIS	0.04 [0.026, 0.054]	7.41E-09	-0.026 [-0.076, 0.023]	2.99E-01	0.066 [0.026, 0.106]	1.39E-03	0.018 [-0.032, 0.067]	4.85E-01
KEGG MEDICUS REFERENCE ACTIVATION OF PRC2.2 BY UBIQUITINATION OF H2AK119 IN GERMLINE GENES	0.027 [0.013, 0.04]	1.13E-04	-0.011 [-0.06, 0.038]	6.62E-01	-0.005 [-0.044, 0.035]	8.25E-01	0.059 [0.01, 0.108]	1.93E-02
WP HEDGEHOG SIGNALING PATHWAY WP4249	0.031 [0.018, 0.045]	5.61E-06	0.053 [0.004, 0.102]	3.54E-02	0.028 [-0.012, 0.068]	1.71E-01	-0.038 [-0.087, 0.012]	1.37E-01
REACTOME CALCITONIN LIKE LIGAND RECEPTORS	0.025 [0.011, 0.038]	3.58E-04	0.05 [0.001, 0.1]	4.52E-02	-0.027 [-0.067, 0.013]	1.80E-01	0.034 [-0.016, 0.083]	1.81E-01
REACTOME PARASITE INFECTION	0.033 [0.02, 0.047]	1.63E-06	-0.031 [-0.08, 0.018]	2.20E-01	0.059 [0.019, 0.1]	3.98E-03	0.003 [-0.046, 0.053]	8.93E-01
KEGG MEDICUS REFERENCE LPA GNA12 13 RHOA SIGNALING PATHWAY	0.036 [0.022, 0.05]	1.95E-07	0 [-0.049, 0.049]	9.93E-01	0.046 [0.006, 0.085]	2.57E-02	-0.035 [-0.084, 0.015]	1.69E-01
WP PARKIN UBIQUITIN PROTEASOMAL SYSTEM PATHWAY	0.028 [0.014, 0.041]	6.52E-05	-0.037 [-0.086, 0.012]	1.42E-01	0.046 [0.006, 0.086]	2.28E-02	0.002 [-0.047, 0.051]	9.40E-01
PID TOLL ENDOGENOUS PATHWAY	0.03 [0.016, 0.044]	1.45E-05	-0.054 [-0.103, -0.005]	3.16E-02	0.021 [-0.019, 0.062]	2.94E-01	0.054 [0.004, 0.103]	3.46E-02

Supplementary Table 9. *Baseline characteristics of cohorts used in clinical outcomes secondary analysis.*

STROKE		
	UKBB (no MRI)	**3C**
N	349,283	5,745
Incident stroke	9,212	232
Ischemic	7,281	184
ICH	1,344	42
Other (sub-arachnoid or undefined)	587	6
Age at baseline assessment (SD)	57.2 (8.0)	74.1 (5.4)
Median Follow-up Time, years [IQR]	13.2 [12.5-13.9]	9.5 [7.0-10.7]
Female (%)	190,137 (54.4)	3,529 (61.4)

DEMENTIA			
	UKBB (no MRI)	**3C**	**Memento**
N	355,180	5,883	2,064
Incident dementia	7,924	662	287
AD	2,728	459	197
VMD	1,187	131	37
Other (undefined or other causes)	4,009	72	53
Age at baseline assessment (SD)	57.2 (8.0)	74.2 (5.4)	70.9 (8.6)
Median Follow-up Time, years [IQR]	13.8 [13.0-14.5]	8.7 [4.0-11.0]	5.0 [2.9-5.1]
Female (%)	192,558 (54.2)	3,571 (60.7)	1,270 (61.5)

Supplementary Table 10. *Associations of WMH-(ps)PRS with stroke outcomes in UKBB.*

	Any Stroke				Ischemic Stroke				Intracerebral Hemorrhage			
PRS	**HR**	**CI_LOW**	**CI_HIGH**	**P**	**HR**	**CI_LOW**	**CI_HIGH**	**P**	**HR**	**CI_LOW**	**CI_HIGH**	**P**
BIOCARTA_NPP1_PATHWAY	1.037	1.016	1.059	0.0004	1.034	1.011	1.058	0.0038	1.071	1.015	1.13	0.0121
Carbohydrate_metabolism	1.004	0.984	1.025	0.6834	1.003	0.98	1.026	0.8183	1.004	0.951	1.059	0.8948
Immune_response_to_viral_infections	1.017	0.997	1.038	0.0981	1.022	0.999	1.046	0.0594	1.014	0.961	1.07	0.6146
KEGG_AMINO_SUGAR_AND_NUCLEOTIDE_SUGAR_METABOLISM	1.019	0.998	1.04	0.0772	1.024	1.001	1.048	0.0429	1.024	0.97	1.08	0.3916
KEGG_MEDICUS_REFERENCE_ACTIVATION_OF_PRC2.2_BY_UBIQUITINATION_OF_H2AK119_IN_GERMLINE_GENES	0.996	0.976	1.017	0.6998	0.99	0.968	1.013	0.4068	0.982	0.931	1.036	0.5025
KEGG_MEDICUS_REFERENCE_LPA_GNA12_13_RHOA_SIGNALING_PATHWAY	1.007	0.986	1.028	0.5118	1.013	0.99	1.037	0.2721	1.017	0.964	1.073	0.5332
KEGG_MEDICUS_REFERENCE_MICROTUBULE_DEPOLYMERIZATION	1.008	0.988	1.029	0.4180	1.015	0.992	1.038	0.2129	0.976	0.925	1.029	0.3635
KEGG_MEDICUS_REFERENCE_REGULATION_OF_GF_RTK_RAS_ERK_SIGNALING_RAS_UBIQUITINATION_BY_CUL3_COMPLEX	0.984	0.964	1.004	0.1114	0.983	0.961	1.006	0.1442	1.019	0.966	1.074	0.4977
KEGG_PEROXISOME	1.001	0.98	1.021	0.9611	1.002	0.979	1.025	0.8971	0.977	0.926	1.031	0.3986
KEGG_SMALL_CELL_LUNG_CANCER	1.024	1.003	1.045	0.0235	1.035	1.011	1.059	0.0034	1.013	0.961	1.069	0.6234
Metabolism_of_nucleotide_sugars_and_glycolysis	1.024	1.003	1.045	0.0250	1.018	0.995	1.042	0.1190	1.072	1.016	1.131	0.0105
Organelle_biosynthesis	1.025	1.005	1.047	0.0161	1.024	1	1.047	0.0478	1.039	0.985	1.096	0.1634
PID_AVB3_OPN_PATHWAY	1.013	0.992	1.034	0.2233	1.009	0.986	1.032	0.4611	1.007	0.954	1.062	0.8064
PID_EPHA_FWDPATHWAY	1.019	0.998	1.04	0.0784	1.02	0.997	1.044	0.0868	1.03	0.977	1.087	0.2738
PID_EPHA2_FWD_PATHWAY	1.005	0.984	1.025	0.6607	1.006	0.983	1.029	0.6188	1.014	0.961	1.07	0.6051
PID_TOLL_ENDOGENOUS_PATHWAY	0.997	0.977	1.018	0.7806	1	0.978	1.024	0.9694	0.971	0.92	1.024	0.2763
Protein_Localization_and_Fatty_acid_metabolism	1.005	0.985	1.026	0.6112	0.999	0.976	1.022	0.9078	1.03	0.977	1.087	0.2725
REACTOME_ALPHA_LINOLENIC_OMEGA3_AND_LINOLEIC_OMEGA6_ACID_METABOLISM	0.986	0.966	1.006	0.1726	0.968	0.946	0.991	0.0056	1.057	1.002	1.115	0.0418
REACTOME_ANCHORING_OF_THE_BASAL_BODY_TO_THE_PLASMA_MEMBRANE	1.018	0.998	1.039	0.0830	1.012	0.989	1.036	0.2904	1.061	1.006	1.119	0.0305
REACTOME_CALCITONIN_LIKE_LIGAND_RECEPTORS	1.026	1.005	1.047	0.0143	1.023	1	1.047	0.0492	1	0.948	1.055	0.9927
REACTOME_CHYLOMICRON_CLEARANCE	1.03	1.01	1.052	0.0040	1.018	0.995	1.042	0.1242	1.122	1.064	1.183	2.00E-05
REACTOME_DISEASES_ASSOCIATED_WITH_GLYCOSAMINOGLYCAN_METABOLISM	0.99	0.97	1.01	0.3236	0.99	0.967	1.013	0.3910	1.005	0.953	1.06	0.8502
REACTOME_ELASTIC_FIBRE_FORMATION	1.007	0.987	1.028	0.5101	1.017	0.994	1.041	0.1447	0.938	0.889	0.99	0.0197
REACTOME_EPH_EPHRIN_SIGNALING	0.999	0.979	1.019	0.9023	1.006	0.983	1.029	0.6118	0.967	0.916	1.02	0.2130
REACTOME_EPHA_MEDIATED_GROWTH_CONE_COLLAPSE	1.005	0.985	1.026	0.6186	1.011	0.988	1.035	0.3339	1.006	0.954	1.062	0.8205
REACTOME_EPIGENETIC_REGULATION_OF_GENE_EXPRESSION	1.02	0.999	1.041	0.0599	1.02	0.997	1.043	0.0939	1.035	0.981	1.092	0.2100
REACTOME_FCGAMMA_RECEPTOR_FCGR_DEPENDENT_PHAGOCYTOSIS	0.997	0.976	1.017	0.7391	1	0.977	1.023	0.9882	0.993	0.941	1.047	0.7885
REACTOME_G0_AND_EARLY_G1	1.01	0.989	1.031	0.3477	1.008	0.985	1.031	0.4987	1.011	0.958	1.066	0.6957
REACTOME_GAP_JUNCTION_TRAFFICKING_AND_REGULATION	1.002	0.982	1.023	0.8425	0.998	0.975	1.021	0.8333	1.011	0.958	1.066	0.6906
REACTOME_HIV_INFECTION	1.016	0.995	1.037	0.1280	1.014	0.991	1.038	0.2321	1.037	0.983	1.094	0.1839
REACTOME_INFECTIOUS_DISEASE	1.017	0.997	1.038	0.0981	1.022	0.999	1.046	0.0594	1.014	0.961	1.07	0.6146
REACTOME_MHC_CLASS_II_ANTIGEN_PRESENTATION	1.008	0.987	1.029	0.4539	1.01	0.987	1.033	0.4093	1.008	0.956	1.064	0.7641
REACTOME_MITOCHONDRIAL_TRANSLATION	0.995	0.975	1.016	0.6493	0.99	0.968	1.013	0.4090	1.015	0.962	1.07	0.5911
REACTOME_MOLECULES_ASSOCIATED_WITH_ELASTIC_FIBRES	1.003	0.983	1.024	0.7768	1.015	0.992	1.039	0.2049	0.93	0.882	0.981	0.0081
REACTOME_ORGANELLE_BIOGENESIS_AND_MAINTENANCE	1.02	1	1.042	0.0523	1.014	0.991	1.038	0.2258	1.064	1.009	1.123	0.0230
REACTOME_PARASITE_INFECTION	0.991	0.971	1.012	0.4028	0.995	0.972	1.018	0.6749	0.999	0.947	1.054	0.9755
REACTOME_PRE_NOTCH_PROCESSING_IN_GOLGI	1.002	0.982	1.023	0.8511	1	0.977	1.023	0.9823	1.006	0.954	1.062	0.8159
REACTOME_PROTEIN_LOCALIZATION	1.014	0.994	1.035	0.1777	1.009	0.986	1.032	0.4637	1.049	0.994	1.107	0.0805
REACTOME_RHO_GTPASE_CYCLE	1.032	1.011	1.053	0.0025	1.039	1.015	1.063	0.0012	1.025	0.972	1.081	0.3649
REACTOME_SIGNALING_BY_PDGF	1.03	1.009	1.051	0.0046	1.038	1.014	1.062	0.0016	1.031	0.977	1.087	0.2686
REACTOME_SIGNALING_BY_RECEPTOR_TYROSINE_KINASES	1.023	1.002	1.044	0.0289	1.03	1.006	1.054	0.0126	1.008	0.955	1.063	0.7720
REACTOME_SIGNALING_BY_RHO_GTPASES_MIRO_GTPASES_AND_RHOBTB3	1.029	1.008	1.05	0.0066	1.035	1.012	1.059	0.0031	1.016	0.963	1.072	0.5504
REACTOME_SIGNALING_BY_VEGF	1.007	0.987	1.028	0.4874	1.012	0.989	1.036	0.3009	1.018	0.965	1.074	0.5209
REACTOME_TCF_DEPENDENT_SIGNALING_IN_RESPONSE_TO_WNT	0.981	0.961	1.002	0.0699	0.984	0.961	1.007	0.1641	0.98	0.929	1.034	0.4533
REACTOME_TRANSCRIPTION_OF_E2F_TARGETS_UNDER_NEGATIVE_CONTROL_BY_DREAM_COMPLEX	1.016	0.995	1.037	0.1330	1.011	0.988	1.035	0.3489	1.026	0.973	1.082	0.3476
REACTOME_VIRAL_INFECTION_PATHWAYS	1.005	0.985	1.026	0.6055	1.006	0.983	1.03	0.5881	1.004	0.952	1.06	0.8750
REACTOME_VITAMIN_B5_PANTOTHENATE_METABOLISM	1	0.979	1.02	0.9642	0.996	0.974	1.02	0.7598	1.024	0.971	1.081	0.3777
Signal_Transduction	1.019	0.999	1.04	0.0659	1.025	1.002	1.049	0.0338	1.004	0.951	1.059	0.8938
WMH_PRS_global	1.06	1.038	1.082	0.0000	1.061	1.037	1.086	0.0000	1.083	1.026	1.143	0.0038
WP_CONGENITAL_GENERALIZED_LIPODYSTROPHY	1.018	0.998	1.039	0.0852	1.022	0.999	1.046	0.0609	1.003	0.951	1.058	0.9129
WP_DISORDERS_OF_GALACTOSE_METABOLISM	1.002	0.982	1.023	0.8491	1.002	0.979	1.025	0.8507	1.01	0.958	1.066	0.7093
WP_EICOSANOID_METABOLISM_VIA_LIPOOXYGENASES_LOX	0.991	0.971	1.011	0.3785	0.988	0.965	1.011	0.2876	1.003	0.951	1.058	0.9160
WP_ESTROGEN_RECEPTOR_PATHWAY	0.989	0.969	1.01	0.2960	0.988	0.966	1.011	0.3111	0.979	0.928	1.033	0.4452
WP_FOCAL_ADHESION	1.012	0.992	1.033	0.2462	1.012	0.989	1.036	0.2954	1.024	0.97	1.08	0.3944
WP_GENES_RELATED_TO_PRIMARY_CILIUM_DEVELOPMENT_BASED_ON_CRISPR	1.018	0.998	1.039	0.0798	1.014	0.991	1.037	0.2429	1.065	1.009	1.123	0.0213
WP_HEDGEHOG_SIGNALING_PATHWAY_WP4249	1.002	0.982	1.023	0.8306	1.002	0.979	1.025	0.8849	1.007	0.954	1.062	0.8106
WP_INTEGRIN_MEDIATED_CELL_ADHESION	1.015	0.994	1.036	0.1657	1.018	0.994	1.041	0.1388	0.993	0.942	1.048	0.8096
WP_OMEGA_3_OMEGA_6_FATTY_ACID_SYNTHESIS	0.991	0.971	1.011	0.3732	0.969	0.947	0.991	0.0070	1.08	1.024	1.139	0.0047
WP_PARKIN_UBIQUITIN_PROTEASOMAL_SYSTEM_PATHWAY	1.002	0.982	1.023	0.8365	1.003	0.98	1.026	0.8139	1.023	0.969	1.079	0.4143
WP_RESISTIN_AS_A_REGULATOR_OF_INFLAMMATION	1.012	0.992	1.033	0.2491	1.004	0.982	1.028	0.7021	1.047	0.992	1.104	0.0960
WP_SPHINGOLIPID_METABOLISM_INTEGRATED_PATHWAY	1.016	0.996	1.037	0.1254	1.027	1.003	1.05	0.0257	0.977	0.926	1.031	0.4003
WP_SPHINGOLIPID_METABOLISM_OVERVIEW	1.016	0.995	1.036	0.1399	1.027	1.004	1.051	0.0235	0.981	0.93	1.035	0.4793

Supplementary Table 11. *WMH-(ps)PRS association with dementia outcomes in UKBB.*

	All-Cause Dementia				Alzheimer's Disease				Vascuar or Mixed Dementia			
PRS	**HR**	**CI_LOW**	**CI_HIGH**	**P**	**HR**	**CI_LOW**	**CI_HIGH**	**P**	**HR**	**CI_LOW**	**CI_HIGH**	**P**
WMH PRS global	1.051	1.028	1.075	1.00E-05	1.018	0.98	1.057	0.35042	1.168	1.103	1.237	0
REACTOME PARASITE INFECTION	1.038	1.015	1.061	0.00093	1.06	1.021	1.1	0.00237	1.022	0.965	1.082	0.458
Metabolism of nucleotide sugars and glycolysis	1.035	1.012	1.058	0.00221	1.033	0.994	1.072	0.09511	1.025	0.968	1.085	0.39498
REACTOME FCGAMMA RECEPTOR FCGR DEPENDENT PHAGOCYTOSIS	1.033	1.011	1.056	0.0035	1.062	1.023	1.102	0.00174	1.02	0.964	1.08	0.49195
WP CONGENITAL GENERALIZED LIPODYSTROPHY	1.033	1.01	1.056	0.00436	1.025	0.987	1.065	0.19339	1.055	0.996	1.117	0.06768
KEGG MEDICUS REFERENCE LPA GNA12 13 RHOA SIGNALING PATHWAY	1.028	1.006	1.051	0.01312	1.012	0.975	1.051	0.5312	1.046	0.988	1.107	0.12505
KEGG AMINO SUGAR AND NUCLEOTIDE SUGAR METABOLISM	1.028	1.006	1.051	0.01428	1.031	0.993	1.07	0.11465	1.033	0.976	1.094	0.25764
REACTOME RHO GTPASE CYCLE	1.026	1.004	1.049	0.02254	1.016	0.979	1.055	0.3938	1.079	1.02	1.143	0.00842
Signal Transduction	1.026	1.003	1.049	0.02381	1.034	0.996	1.074	0.08124	1.015	0.959	1.075	0.61062
WP HEDGEHOG SIGNALING PATHWAY WP4249	1.025	1.003	1.048	0.02689	1.013	0.975	1.051	0.51287	1.059	1.001	1.121	0.04797
REACTOME SIGNALING BY VEGF	1.024	1.001	1.047	0.03639	1.016	0.979	1.055	0.40053	1.04	0.982	1.1	0.18161
PID TOLL ENDOGENOUS PATHWAY	0.978	0.956	1	0.04582	0.98	0.943	1.017	0.28143	0.984	0.93	1.042	0.59014
REACTOME SIGNALING BY RECEPTOR TYROSINE KINASES	1.022	1	1.045	0.04868	1.025	0.987	1.065	0.19603	1.025	0.968	1.085	0.39313
PID EPHA FWDPATHWAY	1.022	0.999	1.044	0.05791	1.014	0.977	1.053	0.45914	0.986	0.932	1.044	0.63476
REACTOME EPH EPHRIN SIGNALING	1.021	0.998	1.043	0.07007	1.011	0.973	1.05	0.57797	1.022	0.965	1.082	0.45764
Immune response to viral infections	1.02	0.998	1.043	0.07828	1.007	0.97	1.046	0.70083	1.042	0.984	1.103	0.15668
REACTOME INFECTIOUS DISEASE	1.02	0.998	1.043	0.07828	1.007	0.97	1.046	0.70083	1.042	0.984	1.103	0.15668
REACTOME MOLECULES ASSOCIATED WITH ELASTIC FIBRES	1.02	0.998	1.043	0.07882	1.018	0.981	1.057	0.34893	1.041	0.984	1.102	0.16392
REACTOME SIGNALING BY RHO GTPASES MIRO GTPASES AND RHOBTB3	1.018	0.996	1.041	0.10444	1.001	0.964	1.039	0.9603	1.079	1.019	1.142	0.00881
REACTOME ELASTIC FIBRE FORMATION	1.018	0.996	1.041	0.10674	1.011	0.974	1.05	0.56448	1.04	0.983	1.101	0.17293
Organelle biosynthesis	1.017	0.995	1.04	0.12652	0.977	0.941	1.014	0.21886	1.085	1.025	1.149	0.00502
REACTOME DISEASES ASSOCIATED WITH GLYCOSAMINOGLYCAN METABOLISM	1.016	0.994	1.038	0.16648	1.03	0.992	1.07	0.12458	0.985	0.93	1.043	0.60183
WP EICOSANOID METABOLISM VIA LIPOOXYGENASES LOX	0.985	0.964	1.007	0.18034	0.988	0.952	1.026	0.53841	0.999	0.944	1.057	0.97214
REACTOME SIGNALING BY PDGF	1.015	0.993	1.037	0.19461	0.995	0.958	1.033	0.79015	1.04	0.982	1.101	0.17747
BIOCARTA NPP1 PATHWAY	1.012	0.99	1.035	0.28339	1.003	0.966	1.041	0.89295	1.047	0.989	1.108	0.11379
WP OMEGA 3 OMEGA 6 FATTY ACID SYNTHESIS	0.989	0.967	1.011	0.30514	0.974	0.938	1.011	0.1683	0.942	0.89	0.997	0.03958
REACTOME TRANSCRIPTION OF E2F TARGETS UNDER NEGATIVE CONTROL BY DREAM COMPLEX	0.989	0.968	1.011	0.3284	0.979	0.943	1.016	0.26134	1.032	0.975	1.093	0.27727
REACTOME EPIGENETIC REGULATION OF GENE EXPRESSION	1.011	0.989	1.033	0.33172	1	0.963	1.038	0.98364	1.033	0.976	1.093	0.26297
Protein Localization and Fatty acid metabolism	0.99	0.968	1.012	0.37802	0.973	0.937	1.01	0.15541	1.009	0.953	1.068	0.76995
REACTOME MITOCHONDRIAL TRANSLATION	1.009	0.987	1.032	0.41203	0.992	0.955	1.029	0.65829	0.99	0.935	1.048	0.72296
KEGG MEDICUS REFERENCE MICROTUBULE DEPOLYMERIZATION	0.991	0.969	1.013	0.41958	1.004	0.967	1.042	0.83653	1.019	0.962	1.078	0.52084
REACTOME CALCITONIN LIKE LIGAND RECEPTORS	0.991	0.969	1.013	0.42003	0.999	0.962	1.037	0.94977	0.996	0.941	1.055	0.9025
REACTOME EPHA MEDIATED GROWTH CONE COLLAPSE	1.009	0.987	1.031	0.4376	0.994	0.958	1.033	0.76875	1.014	0.958	1.074	0.62505
REACTOME VIRAL INFECTION PATHWAYS	1.008	0.986	1.031	0.45679	0.992	0.956	1.03	0.68822	1.031	0.974	1.092	0.28716
WP FOCAL ADHESION	1.008	0.986	1.031	0.46313	1	0.963	1.038	0.99569	0.998	0.943	1.057	0.94382
REACTOME ANCHORING OF THE BASAL BODY TO THE PLASMA MEMBRANE	0.992	0.97	1.014	0.46513	0.953	0.918	0.99	0.01267	1.077	1.018	1.14	0.01022
WP DISORDERS OF GALACTOSE METABOLISM	1.008	0.986	1.03	0.4728	0.985	0.949	1.022	0.42406	1.043	0.986	1.104	0.13974
PID EPHA2 FWD PATHWAY	1.007	0.985	1.03	0.51406	0.999	0.962	1.037	0.94364	0.971	0.917	1.028	0.3123
WP RESISTIN AS A REGULATOR OF INFLAMMATION	0.993	0.971	1.015	0.51963	1.001	0.964	1.04	0.94625	1.004	0.948	1.063	0.89208
Carbohydrate metabolism	1.007	0.985	1.03	0.5258	0.982	0.946	1.02	0.34686	1.047	0.989	1.109	0.11109
REACTOME CHYLOMICRON CLEARANCE	1.008	0.983	1.034	0.53608	1.011	0.969	1.056	0.60394	1.026	0.961	1.095	0.44478
REACTOME TCF DEPENDENT SIGNALING IN RESPONSE TO WNT	1.007	0.985	1.029	0.53916	0.99	0.954	1.028	0.6028	1.015	0.959	1.075	0.60565
REACTOME GAP JUNCTION TRAFFICKING AND REGULATION	0.993	0.972	1.015	0.54036	0.992	0.956	1.03	0.67602	0.997	0.942	1.055	0.91757
KEGG MEDICUS REFERENCE ACTIVATION OF PRC2.2 BY UBIQUITINATION OF H2AK119 IN GERMLINE GENES	1.007	0.985	1.029	0.55063	0.992	0.955	1.03	0.66524	1.03	0.973	1.09	0.30247
PID AVB3 OPN PATHWAY	1.007	0.985	1.029	0.55199	1.004	0.967	1.042	0.84646	0.983	0.928	1.04	0.54741
REACTOME PROTEIN LOCALIZATION	1.007	0.984	1.029	0.56042	1.001	0.964	1.04	0.94913	1.038	0.98	1.1	0.1994
KEGG MEDICUS REFERENCE REGULATION OF GF RTK RAS ERK SIGNALING RAS UBIQUITINATION BY CUL3 COMPLEX	1.007	0.985	1.029	0.56059	0.972	0.936	1.009	0.13312	1.047	0.99	1.108	0.1088
REACTOME PRE NOTCH PROCESSING IN GOLGI	1.006	0.984	1.029	0.57392	0.997	0.96	1.035	0.87884	1.047	0.989	1.108	0.11578
KEGG PEROXISOME	0.994	0.972	1.016	0.57432	0.976	0.94	1.013	0.20253	0.987	0.933	1.045	0.65785
WP ESTROGEN RECEPTOR PATHWAY	1.006	0.984	1.028	0.59977	1.025	0.987	1.064	0.20048	0.949	0.896	1.004	0.07089
REACTOME G0 AND EARLY G1	1.005	0.983	1.028	0.63549	0.987	0.951	1.025	0.49015	1.051	0.993	1.113	0.08444
WP INTEGRIN MEDIATED CELL ADHESION	1.004	0.982	1.027	0.71106	0.981	0.944	1.018	0.30554	0.979	0.925	1.037	0.47372
WP GENES RELATED TO PRIMARY CILIUM DEVELOPMENT BASED ON CRISPR	0.996	0.974	1.018	0.71275	0.981	0.945	1.018	0.30826	1.037	0.98	1.098	0.20701
REACTOME ALPHA LINOLENIC OMEGA3 AND LINOLEIC OMEGA6 ACID METABOLISM	0.997	0.975	1.019	0.76892	0.991	0.955	1.029	0.65132	0.937	0.885	0.992	0.02455
REACTOME VITAMIN B5 PANTOTHENATE METABOLISM	1.003	0.981	1.025	0.78712	1.006	0.968	1.044	0.77132	0.993	0.938	1.051	0.81408
WP PARKIN UBIQUITIN PROTEASOMAL SYSTEM PATHWAY	0.997	0.975	1.019	0.79786	0.988	0.952	1.026	0.54049	1.034	0.977	1.095	0.24604
WP SPHINGOLIPID METABOLISM OVERVIEW	0.997	0.976	1.019	0.80149	0.982	0.946	1.019	0.33337	1.001	0.946	1.06	0.96752
REACTOME HIV INFECTION	0.998	0.976	1.02	0.86144	0.973	0.937	1.01	0.15109	1.037	0.98	1.098	0.21209
REACTOME ORGANELLE BIOGENESIS AND MAINTENANCE	1.002	0.98	1.024	0.87695	0.979	0.942	1.016	0.25819	1.065	1.006	1.128	0.03041
REACTOME MHC CLASS II ANTIGEN PRESENTATION	1.002	0.98	1.024	0.88675	0.999	0.962	1.037	0.93965	1.019	0.962	1.079	0.51956
KEGG SMALL CELL LUNG CANCER	1.002	0.98	1.024	0.89234	0.984	0.947	1.022	0.39538	1.016	0.96	1.076	0.58744
WP SPHINGOLIPID METABOLISM INTEGRATED PATHWAY	0.999	0.977	1.021	0.92811	0.984	0.948	1.022	0.41206	0.997	0.942	1.056	0.9293

Supplementary Table 12. *WMH-(ps)PRS association with stroke outcomes in 3C.*

	Any Stroke				Ischemic Stroke				Intracerebral Hemorrhage			
PRS	**HR**	**CI_LOW**	**CI_HIGH**	**P**	**HR**	**CI_LOW**	**CI_HIGH**	**P**	**HR**	**CI_LOW**	**CI_HIGH**	**P**
BIOCARTA NPP1 PATHWAY	1.088	0.957	1.237	0.198	1.04	0.9	1.202	0.595	1.185	0.876	1.602	0.270
Carbohydrate metabolism	1.058	0.931	1.201	0.387	1.046	0.907	1.206	0.539	1.198	0.886	1.62	0.240
Immune response to viral infections	0.967	0.85	1.101	0.617	0.93	0.804	1.075	0.325	1.185	0.872	1.61	0.279
KEGG AMINO SUGAR AND NUCLEOTIDE SUGAR METABOLISM	0.964	0.847	1.097	0.578	0.963	0.833	1.112	0.607	1.012	0.746	1.373	0.938
KEGG MEDICUS REFERENCE ACTIVATION OF PRC2.2 BY UBIQUITINATION OF H2AK119 IN GERMLINE GENES	0.996	0.875	1.133	0.948	0.985	0.851	1.139	0.834	1.153	0.854	1.557	0.352
KEGG MEDICUS REFERENCE LPA GNA12 13 RHOA SIGNALING PATHWAY	1.053	0.923	1.2	0.443	1.041	0.899	1.206	0.590	1.134	0.834	1.542	0.422
KEGG MEDICUS REFERENCE MICROTUBULE DEPOLYMERIZATION	1.137	0.999	1.294	0.051	1.168	1.01	1.351	0.036	1.087	0.804	1.469	0.587
KEGG MEDICUS REFERENCE REGULATION OF GF RTK RAS ERK SIGNALING RAS UBIQUITINATION BY CUL3 COMPLEX	1.051	0.924	1.195	0.449	1.009	0.873	1.166	0.908	1.284	0.954	1.73	0.099
KEGG PEROXISOME	0.967	0.849	1.103	0.620	0.967	0.835	1.12	0.658	0.977	0.716	1.335	0.885
KEGG SMALL CELL LUNG CANCER	0.975	0.857	1.11	0.704	0.906	0.784	1.047	0.183	1.112	0.819	1.51	0.497
Metabolism of nucleotide sugars and glycolysis	0.97	0.853	1.103	0.642	0.924	0.799	1.067	0.280	1.222	0.906	1.649	0.190
Organelle biosynthesis	1.036	0.911	1.178	0.591	1.045	0.904	1.207	0.554	1	0.743	1.345	0.998
PID AVB3 OPN PATHWAY	1.025	0.901	1.165	0.711	1.011	0.875	1.167	0.886	1.172	0.862	1.592	0.311
PID EPHA FWDPATHWAY	1.088	0.956	1.238	0.203	1.096	0.948	1.268	0.217	1.143	0.842	1.552	0.393
PID EPHA2 FWD PATHWAY	1.071	0.942	1.219	0.296	1.034	0.895	1.194	0.654	1.388	1.017	1.893	0.039
PID TOLL ENDOGENOUS PATHWAY	1.091	0.96	1.239	0.184	1.092	0.946	1.261	0.230	1.08	0.801	1.456	0.613
Protein Localization and Fatty acid metabolism	0.943	0.828	1.075	0.383	0.928	0.801	1.075	0.320	0.994	0.73	1.355	0.970
REACTOME ALPHA LINOLENIC OMEGA3 AND LINOLEIC OMEGA6 ACID METABOLISM	1.047	0.921	1.19	0.485	1.055	0.914	1.219	0.463	0.927	0.683	1.259	0.628
REACTOME ANCHORING OF THE BASAL BODY TO THE PLASMA MEMBRANE	1.066	0.936	1.213	0.334	1.06	0.916	1.226	0.434	1.037	0.766	1.402	0.816
REACTOME CALCITONIN LIKE LIGAND RECEPTORS	1.081	0.951	1.229	0.234	1.09	0.943	1.258	0.243	0.993	0.734	1.344	0.963
REACTOME CHYLOMICRON CLEARANCE	1.081	0.951	1.23	0.234	1.111	0.962	1.283	0.151	0.965	0.707	1.316	0.821
REACTOME DISEASES ASSOCIATED WITH GLYCOSAMINOGLYCAN METABOLISM	1.036	0.911	1.178	0.594	1.042	0.902	1.204	0.578	1.077	0.798	1.454	0.627
REACTOME ELASTIC FIBRE FORMATION	1.059	0.931	1.205	0.381	1.092	0.945	1.261	0.233	0.906	0.668	1.228	0.525
REACTOME EPH EPHRIN SIGNALING	1.141	1.002	1.3	0.047	1.11	0.959	1.285	0.163	1.236	0.912	1.677	0.172
REACTOME EPHA MEDIATED GROWTH CONE COLLAPSE	1.135	0.996	1.293	0.058	1.103	0.953	1.277	0.190	1.352	0.997	1.834	0.052
REACTOME EPIGENETIC REGULATION OF GENE EXPRESSION	0.921	0.809	1.047	0.208	0.9	0.779	1.041	0.155	1.013	0.747	1.373	0.933
REACTOME FCGAMMA RECEPTOR FCGR DEPENDENT PHAGOCYTOSIS	0.931	0.817	1.06	0.280	0.936	0.809	1.084	0.378	0.936	0.687	1.275	0.676
REACTOME G0 AND EARLY G1	1.067	0.937	1.214	0.327	1.097	0.949	1.268	0.210	0.95	0.702	1.287	0.741
REACTOME GAP JUNCTION TRAFFICKING AND REGULATION	1.078	0.947	1.227	0.258	1.082	0.936	1.252	0.286	1.082	0.797	1.47	0.612
REACTOME HIV INFECTION	0.931	0.817	1.06	0.278	0.899	0.777	1.041	0.155	1.045	0.771	1.418	0.775
REACTOME INFECTIOUS DISEASE	0.967	0.85	1.101	0.617	0.93	0.804	1.075	0.325	1.185	0.872	1.61	0.279
REACTOME MHC CLASS II ANTIGEN PRESENTATION	1.098	0.964	1.25	0.158	1.045	0.903	1.208	0.557	1.324	0.973	1.802	0.074
REACTOME MITOCHONDRIAL TRANSLATION	1.061	0.933	1.207	0.364	1.04	0.899	1.202	0.599	1.089	0.803	1.475	0.584
REACTOME MOLECULES ASSOCIATED WITH ELASTIC FIBRES	1.075	0.945	1.223	0.273	1.12	0.969	1.294	0.125	0.871	0.642	1.182	0.376
REACTOME ORGANELLE BIOGENESIS AND MAINTENANCE	1.015	0.892	1.154	0.824	1.033	0.894	1.194	0.658	0.955	0.707	1.29	0.764
REACTOME PARASITE INFECTION	0.963	0.845	1.096	0.565	0.968	0.837	1.12	0.663	0.967	0.71	1.318	0.832
REACTOME PRE NOTCH PROCESSING IN GOLGI	0.839	0.737	0.954	0.008	0.824	0.713	0.953	0.009	0.964	0.711	1.306	0.812
REACTOME PROTEIN LOCALIZATION	0.887	0.779	1.01	0.071	0.853	0.737	0.986	0.032	0.972	0.714	1.323	0.855
REACTOME RHO GTPASE CYCLE	1.005	0.884	1.143	0.938	0.985	0.853	1.138	0.842	1.079	0.794	1.464	0.628
REACTOME SIGNALING BY PDGF	1.077	0.946	1.225	0.262	1.073	0.929	1.241	0.337	1.151	0.847	1.565	0.368
REACTOME SIGNALING BY RECEPTOR TYROSINE KINASES	0.999	0.877	1.137	0.983	0.979	0.847	1.133	0.778	1.135	0.838	1.539	0.413
REACTOME SIGNALING BY RHO GTPASES MIRO GTPASES AND RHOBTB3	1.059	0.932	1.204	0.381	1.026	0.889	1.185	0.727	1.235	0.91	1.674	0.175
REACTOME SIGNALING BY VEGF	0.992	0.872	1.129	0.906	0.993	0.859	1.148	0.923	0.971	0.718	1.313	0.847
REACTOME TCF DEPENDENT SIGNALING IN RESPONSE TO WNT	0.991	0.871	1.128	0.894	1.026	0.888	1.187	0.727	0.831	0.612	1.129	0.237
REACTOME TRANSCRIPTION OF E2F TARGETS UNDER NEGATIVE CONTROL BY DREAM COMPLEX	1.036	0.91	1.179	0.591	1.066	0.922	1.233	0.387	0.945	0.698	1.279	0.713
REACTOME VIRAL INFECTION PATHWAYS	1.01	0.888	1.149	0.882	0.96	0.831	1.11	0.582	1.298	0.959	1.757	0.091
REACTOME VITAMIN B5 PANTOTHENATE METABOLISM	1.006	0.885	1.143	0.930	0.992	0.859	1.146	0.914	1.12	0.827	1.516	0.464
Signal Transduction	0.98	0.86	1.116	0.761	0.959	0.829	1.11	0.577	1.14	0.839	1.548	0.401
WMH PRSCS global	1.047	0.919	1.194	0.489	1.053	0.909	1.22	0.488	1.082	0.796	1.472	0.614
WP CONGENITAL GENERALIZED LIPODYSTROPHY	1.102	0.968	1.254	0.142	1.012	0.876	1.17	0.868	1.421	1.045	1.933	0.025
WP DISORDERS OF GALACTOSE METABOLISM	1.046	0.92	1.189	0.494	1.023	0.885	1.182	0.759	1.022	0.755	1.383	0.888
WP EICOSANOID METABOLISM VIA LIPOOXYGENASES LOX	1.05	0.922	1.196	0.461	1.081	0.934	1.251	0.299	0.964	0.708	1.312	0.816
WP ESTROGEN RECEPTOR PATHWAY	0.94	0.826	1.071	0.352	0.94	0.813	1.087	0.404	0.944	0.697	1.28	0.712
WP FOCAL ADHESION	1.082	0.951	1.231	0.232	1.074	0.929	1.242	0.334	1.082	0.799	1.467	0.610
WP GENES RELATED TO PRIMARY CILIUM DEVELOPMENT BASED ON CRISPR	1.048	0.921	1.192	0.476	1.067	0.923	1.233	0.381	1.012	0.75	1.366	0.937
WP HEDGEHOG SIGNALING PATHWAY WP4249	0.987	0.868	1.123	0.841	0.967	0.837	1.118	0.654	1.007	0.746	1.36	0.963
WP INTEGRIN MEDIATED CELL ADHESION	1.121	0.985	1.276	0.084	1.115	0.965	1.29	0.140	1.14	0.843	1.54	0.394
WP OMEGA 3 OMEGA 6 FATTY ACID SYNTHESIS	1.004	0.882	1.144	0.946	1.007	0.871	1.165	0.922	1.044	0.769	1.416	0.784
WP PARKIN UBIQUITIN PROTEASOMAL SYSTEM PATHWAY	1.041	0.916	1.183	0.541	1.019	0.882	1.177	0.798	1.062	0.786	1.434	0.697
WP RESISTIN AS A REGULATOR OF INFLAMMATION	0.96	0.845	1.091	0.534	0.995	0.862	1.148	0.942	0.752	0.558	1.014	0.061
WP SPHINGOLIPID METABOLISM INTEGRATED PATHWAY	0.998	0.878	1.135	0.976	1.006	0.871	1.162	0.934	0.876	0.65	1.181	0.385
WP SPHINGOLIPID METABOLISM OVERVIEW	1.006	0.885	1.143	0.930	1.014	0.878	1.171	0.849	0.895	0.664	1.207	0.467

Supplementary Table 13. *WMH-(ps)PRS association with dementia outcomes in 3C.*

	ACD				AD				VMD			
PRS	**HR**	**CI_LOW**	**CI_HIGH**	**P**	**HR**	**CI_LOW**	**CI_HIGH**	**P**	**HR**	**CI_LOW**	**CI_HIGH**	**P**
Immune response to viral infections	1.043	0.965	1.126	0.28844	1.117	1.018	1.226	0.0191	0.873	0.734	1.038	0.12443
REACTOME INFECTIOUS DISEASE	1.043	0.965	1.126	0.28844	1.117	1.018	1.226	0.0191	0.873	0.734	1.038	0.12443
REACTOME MHC CLASS II ANTIGEN PRESENTATION	1.029	0.953	1.11	0.4716	1.106	1.009	1.212	0.03199	0.91	0.766	1.081	0.28116
REACTOME VIRAL INFECTION PATHWAYS	1.048	0.97	1.132	0.23076	1.099	1.002	1.206	0.04578	0.907	0.763	1.079	0.27053
WP SPHINGOLIPID METABOLISM INTEGRATED PATHWAY	1.045	0.969	1.127	0.25031	1.081	0.988	1.183	0.09066	0.847	0.715	1.004	0.05588
REACTOME PARASITE INFECTION	0.928	0.859	1.003	0.05987	0.925	0.843	1.015	0.09933	0.916	0.769	1.092	0.32697
WP SPHINGOLIPID METABOLISM OVERVIEW	1.042	0.966	1.124	0.28303	1.077	0.984	1.179	0.10623	0.87	0.734	1.031	0.10827
PID AVB3 OPN PATHWAY	1.078	0.999	1.163	0.05416	1.078	0.983	1.181	0.10999	1.096	0.924	1.299	0.29347
Organelle biosynthesis	1.039	0.963	1.12	0.32529	1.066	0.974	1.168	0.16494	0.937	0.792	1.11	0.45143
REACTOME ORGANELLE BIOGENESIS AND MAINTENANCE	1.035	0.959	1.116	0.37872	1.065	0.973	1.167	0.17374	0.94	0.794	1.115	0.47893
WP RESISTIN AS A REGULATOR OF INFLAMMATION	1.04	0.963	1.123	0.32261	1.066	0.972	1.17	0.17482	1.024	0.862	1.216	0.78595
WP ESTROGEN RECEPTOR PATHWAY	0.958	0.887	1.035	0.27934	0.94	0.856	1.032	0.19166	1.041	0.874	1.239	0.65466
KEGG MEDICUS REFERENCE MICROTUBULE DEPOLYMERIZATION	1.038	0.962	1.12	0.33454	1.061	0.968	1.162	0.2062	1.076	0.907	1.278	0.40067
Protein Localization and Fatty acid metabolism	1.027	0.95	1.11	0.50527	1.057	0.963	1.161	0.24547	0.959	0.806	1.141	0.636
WP CONGENITAL GENERALIZED LIPODYSTROPHY	0.969	0.898	1.045	0.41114	0.949	0.867	1.039	0.25832	0.988	0.833	1.171	0.8875
WP EICOSANOID METABOLISM VIA LIPOOXYGENASES LOX	0.972	0.899	1.05	0.46548	0.95	0.866	1.043	0.28594	1.009	0.849	1.199	0.9169
WP HEDGEHOG SIGNALING PATHWAY WP4249	0.975	0.904	1.052	0.51355	0.954	0.871	1.045	0.31099	1.031	0.87	1.221	0.72477
BIOCARTA NPP1 PATHWAY	1.046	0.968	1.13	0.25162	1.047	0.954	1.149	0.33125	1.114	0.936	1.327	0.22298
REACTOME SIGNALING BY PDGF	1.032	0.955	1.115	0.42796	1.047	0.954	1.149	0.33567	1.006	0.846	1.196	0.94846
REACTOME SIGNALING BY RECEPTOR TYROSINE KINASES	0.999	0.925	1.079	0.97793	1.045	0.953	1.147	0.34818	0.899	0.755	1.07	0.23026
KEGG MEDICUS REFERENCE ACTIVATION OF PRC2.2 BY UBIQUITINATION OF H2AK119 IN GERMLINE GENES	0.94	0.871	1.016	0.11844	0.957	0.872	1.049	0.34865	1.036	0.871	1.232	0.69131
PID TOLL ENDOGENOUS PATHWAY	0.962	0.891	1.039	0.32481	0.958	0.874	1.05	0.3571	0.963	0.81	1.146	0.67376
Signal Transduction	0.991	0.918	1.071	0.82478	1.04	0.948	1.141	0.40803	0.885	0.744	1.054	0.16989
WP GENES RELATED TO PRIMARY CILIUM DEVELOPMENT BASED ON CRISPR	1.055	0.977	1.138	0.17072	1.037	0.947	1.136	0.43294	1.048	0.883	1.244	0.58963
REACTOME CHYLOMICRON CLEARANCE	1.016	0.935	1.103	0.7132	1.04	0.942	1.148	0.43613	0.98	0.814	1.18	0.83174
REACTOME FCGAMMA RECEPTOR FCGR DEPENDENT PHAGOCYTOSIS	0.95	0.881	1.026	0.19188	0.965	0.88	1.057	0.43976	0.908	0.764	1.078	0.27102
REACTOME PROTEIN LOCALIZATION	0.996	0.922	1.075	0.91158	1.034	0.943	1.134	0.47683	0.941	0.791	1.119	0.48973
REACTOME DISEASES ASSOCIATED WITH GLYCOSAMINOGLYCAN METABOLISM	0.995	0.921	1.075	0.9004	1.034	0.942	1.134	0.48296	0.945	0.794	1.126	0.52807
WMH PRSCS global	1.032	0.955	1.115	0.42852	1.034	0.942	1.134	0.48507	1.1	0.924	1.311	0.28406
WP FOCAL ADHESION	1.034	0.958	1.117	0.39265	1.033	0.942	1.134	0.48876	1.08	0.91	1.283	0.37899
REACTOME EPIGENETIC REGULATION OF GENE EXPRESSION	1.008	0.934	1.087	0.84608	1.032	0.942	1.13	0.49878	0.96	0.81	1.137	0.6377
REACTOME VITAMIN B5 PANTOTHENATE METABOLISM	1.019	0.945	1.1	0.61991	1.031	0.941	1.13	0.5101	1.024	0.862	1.216	0.78734
KEGG MEDICUS REFERENCE LPA GNA12 13 RHOA SIGNALING PATHWAY	1.018	0.942	1.099	0.65816	0.972	0.885	1.066	0.54443	1.051	0.883	1.252	0.57618
KEGG PEROXISOME	0.996	0.922	1.075	0.91161	1.028	0.938	1.127	0.55691	0.954	0.803	1.133	0.59345
KEGG AMINO SUGAR AND NUCLEOTIDE SUGAR METABOLISM	0.976	0.903	1.054	0.52757	0.973	0.887	1.067	0.56007	0.875	0.736	1.04	0.12953
REACTOME MOLECULES ASSOCIATED WITH ELASTIC FIBRES	0.973	0.901	1.05	0.47662	0.974	0.889	1.068	0.57921	1.004	0.846	1.19	0.9658
REACTOME ELASTIC FIBRE FORMATION	0.981	0.909	1.059	0.62324	0.977	0.891	1.071	0.6162	1.014	0.855	1.202	0.87491
REACTOME EPH EPHRIN SIGNALING	1.024	0.948	1.105	0.54978	1.024	0.934	1.122	0.62006	1.029	0.866	1.222	0.74847
KEGG SMALL CELL LUNG CANCER	1.027	0.95	1.11	0.50334	1.024	0.932	1.125	0.62123	1.114	0.935	1.328	0.22808
REACTOME SIGNALING BY VEGF	0.967	0.896	1.044	0.3956	0.979	0.893	1.074	0.65429	0.908	0.764	1.079	0.27415
KEGG MEDICUS REFERENCE REGULATION OF GF RTK RAS ERK SIGNALING RAS UBIQUITINATION BY CUL3 COMPLEX	0.972	0.899	1.05	0.47202	0.979	0.892	1.075	0.65778	0.903	0.758	1.075	0.25064
WP DISORDERS OF GALACTOSE METABOLISM	0.978	0.905	1.056	0.56143	0.979	0.893	1.074	0.65892	0.943	0.792	1.123	0.51145
REACTOME EPHA MEDIATED GROWTH CONE COLLAPSE	1.021	0.945	1.103	0.6037	1.021	0.93	1.12	0.6678	0.959	0.805	1.142	0.63695
REACTOME PRE NOTCH PROCESSING IN GOLGI	0.963	0.893	1.039	0.33641	0.981	0.896	1.075	0.68213	0.935	0.787	1.11	0.44208
REACTOME G0 AND EARLY G1	0.966	0.895	1.042	0.37073	0.981	0.895	1.075	0.6833	0.976	0.822	1.158	0.78035
Carbohydrate metabolism	0.991	0.919	1.069	0.81442	1.019	0.93	1.116	0.69134	0.951	0.802	1.127	0.55956
REACTOME MITOCHONDRIAL TRANSLATION	1.019	0.944	1.099	0.63199	0.982	0.896	1.076	0.69766	1.019	0.858	1.21	0.83043
WP OMEGA 3 OMEGA 6 FATTY ACID SYNTHESIS	0.985	0.913	1.063	0.69924	0.983	0.897	1.077	0.71104	0.958	0.807	1.138	0.62706
REACTOME ALPHA LINOLENIC OMEGA3 AND LINOLEIC OMEGA6 ACID METABOLISM	0.988	0.915	1.067	0.75956	1.016	0.926	1.114	0.7421	0.95	0.801	1.128	0.5596
REACTOME CALCITONIN LIKE LIGAND RECEPTORS	0.965	0.894	1.042	0.36876	0.985	0.898	1.08	0.74453	0.975	0.82	1.158	0.76958
REACTOME TCF DEPENDENT SIGNALING IN RESPONSE TO WNT	1.002	0.928	1.081	0.96774	1.014	0.926	1.111	0.75822	0.924	0.779	1.095	0.36237
REACTOME HIV INFECTION	0.966	0.896	1.043	0.38102	0.987	0.901	1.082	0.78264	0.915	0.77	1.087	0.31105
Metabolism of nucleotide sugars and glycolysis	0.959	0.888	1.035	0.28546	0.99	0.903	1.086	0.83306	0.837	0.705	0.993	0.04185
REACTOME ANCHORING OF THE BASAL BODY TO THE PLASMA MEMBRANE	1.007	0.933	1.087	0.85594	1.009	0.921	1.106	0.84052	1.016	0.856	1.206	0.85698
REACTOME SIGNALING BY RHO GTPASES MIRO GTPASES AND RHOBTB3	1.023	0.947	1.104	0.56374	1.009	0.921	1.106	0.84498	1.067	0.898	1.269	0.45868
WP PARKIN UBIQUITIN PROTEASOMAL SYSTEM PATHWAY	1.017	0.944	1.096	0.6592	1.009	0.922	1.104	0.84849	1.019	0.862	1.204	0.82563
REACTOME TRANSCRIPTION OF E2F TARGETS UNDER NEGATIVE CONTROL BY DREAM COMPLEX	0.995	0.921	1.074	0.88847	1.008	0.92	1.106	0.85893	1.02	0.86	1.21	0.81765
REACTOME RHO GTPASE CYCLE	1.022	0.946	1.104	0.5855	1.008	0.918	1.106	0.87097	1.067	0.896	1.271	0.46425
WP INTEGRIN MEDIATED CELL ADHESION	1.004	0.93	1.083	0.9282	1.004	0.917	1.101	0.92598	1.04	0.876	1.234	0.65376
PID EPHA FWDPATHWAY	0.995	0.921	1.075	0.90679	0.999	0.91	1.096	0.97556	1.055	0.885	1.256	0.55138
PID EPHA2 FWD PATHWAY	0.992	0.919	1.071	0.83499	0.999	0.911	1.094	0.97771	1.032	0.869	1.225	0.7197

Supplementary Table 14. *WMH-(ps)PRS association with dementia outcomes in Memento.*

	ACD				AD				VMD			
PRS	**HR**	**CI_LOW**	**CI_HIGH**	**P**	**HR**	**CI_LOW**	**CI_HIGH**	**P**	**HR**	**CI_LOW**	**CI_HIGH**	**P**
WP DISORDERS OF GALACTOSE METABOLISM	1.166	1.039	1.31	0.00924	1.292	1.124	1.485	0.00032	0.927	0.664	1.295	0.65869
Carbohydrate metabolism disorders	1.171	1.038	1.322	0.01032	1.179	1.019	1.365	0.02711	1.199	0.861	1.668	0.2827
REACTOME MITOCHONDRIAL TRANSLATION	0.879	0.78	0.991	0.0343	0.879	0.76	1.016	0.08119	0.964	0.691	1.346	0.83037
REACTOME CALCITONIN LIKE LIGAND RECEPTORS	1.134	1.005	1.279	0.04112	1.079	0.933	1.248	0.3049	1.253	0.898	1.749	0.18353
WP CONGENITAL GENERALIZED LIPODYSTROPHY	1.104	0.979	1.244	0.10542	1.072	0.928	1.239	0.34336	1.046	0.756	1.447	0.7853
PID TOLL ENDOGENOUS PATHWAY	0.912	0.813	1.023	0.1157	0.914	0.796	1.05	0.20544	1.023	0.748	1.401	0.88594
BIOCARTA NPP1 PATHWAY	1.096	0.973	1.235	0.13089	1.154	1	1.333	0.04961	1.054	0.762	1.458	0.75194
KEGG PEROXISOME	1.095	0.968	1.238	0.14816	1.035	0.89	1.203	0.65368	1.613	1.157	2.248	0.00479
REACTOME HIV INFECTION	1.092	0.969	1.229	0.1488	1.127	0.976	1.302	0.10331	1.114	0.804	1.544	0.51709
Protein Localization and Fatty acid metabolism	1.092	0.968	1.233	0.15358	1.017	0.879	1.177	0.81951	1.441	1.024	2.028	0.03592
KEGG AMINO SUGAR AND NUCLEOTIDE SUGAR METABOLISM	1.081	0.963	1.214	0.18762	1.091	0.949	1.254	0.22186	1.125	0.808	1.567	0.48437
Metabolism of nucleotide sugars and glycolysis	1.081	0.963	1.213	0.18889	1.088	0.947	1.249	0.23184	1.053	0.749	1.48	0.76788
KEGG MEDICUS REFERENCE REGULATION OF GF RTK RAS ERK SIGNALING RAS UBIQUITINATION BY CUL3 COMPLEX	1.077	0.959	1.21	0.21002	1.074	0.934	1.235	0.31815	1.129	0.825	1.545	0.44802
REACTOME EPHA MEDIATED GROWTH CONE COLLAPSE	0.928	0.826	1.043	0.21211	0.918	0.797	1.056	0.22974	1.288	0.934	1.774	0.12224
WP SPHINGOLIPID METABOLISM OVERVIEW	1.08	0.956	1.22	0.21819	1.198	1.032	1.39	0.01754	0.932	0.667	1.301	0.67909
REACTOME EPH EPHRIN SIGNALING	0.93	0.825	1.049	0.2391	0.87	0.753	1.005	0.05837	1.123	0.807	1.563	0.4915
PID AVB3 OPN PATHWAY	0.933	0.829	1.049	0.24405	0.99	0.859	1.142	0.89325	0.772	0.562	1.06	0.1093
Signal Transduction	1.07	0.949	1.206	0.26714	1.119	0.965	1.296	0.13596	1.183	0.855	1.637	0.31013
Immune response to viral infections	1.07	0.95	1.205	0.26774	1.146	0.994	1.322	0.06084	0.925	0.672	1.273	0.63085
REACTOME INFECTIOUS DISEASE	1.07	0.95	1.205	0.26774	1.146	0.994	1.322	0.06084	0.925	0.672	1.273	0.63085
REACTOME DISEASES ASSOCIATED WITH GLYCOSAMINOGLYCAN METABOLISM	1.067	0.95	1.199	0.27396	1.018	0.885	1.171	0.80162	1.056	0.764	1.46	0.74183
WP SPHINGOLIPID METABOLISM INTEGRATED PATHWAY	1.068	0.944	1.209	0.29552	1.195	1.027	1.39	0.02089	0.924	0.661	1.292	0.64392
REACTOME VITAMIN B5 PANTOTHENATE METABOLISM	1.064	0.945	1.197	0.30445	1.011	0.875	1.169	0.87843	1.338	0.965	1.855	0.08079
REACTOME PARASITE INFECTION	1.062	0.944	1.194	0.31577	1.117	0.97	1.285	0.1247	0.975	0.701	1.358	0.88221
REACTOME ALPHA LINOLENIC OMEGA3 AND LINOLEIC OMEGA6 ACID METABOLISM	1.056	0.941	1.187	0.35439	1.107	0.963	1.274	0.15342	1.162	0.839	1.609	0.36664
REACTOME SIGNALING BY VEGF	1.057	0.94	1.188	0.35622	1.128	0.98	1.299	0.0943	0.99	0.707	1.387	0.95523
REACTOME PRE NOTCH PROCESSING IN GOLGI	1.056	0.94	1.186	0.35641	1.048	0.911	1.205	0.51213	1.002	0.731	1.373	0.9899
REACTOME SIGNALING BY RECEPTOR TYROSINE KINASES	1.048	0.931	1.181	0.43619	1.079	0.932	1.248	0.3097	1.165	0.843	1.61	0.35359
KEGG MEDICUS REFERENCE MICROTUBULE DEPOLYMERIZATION	1.047	0.929	1.18	0.44912	0.999	0.864	1.155	0.98895	1.131	0.82	1.559	0.45358
REACTOME TRANSCRIPTION OF E2F TARGETS UNDER NEGATIVE CONTROL BY DREAM COMPLEX	0.953	0.839	1.083	0.4615	0.948	0.814	1.105	0.49522	1.034	0.715	1.497	0.85784
WP HEDGEHOG SIGNALING PATHWAY WP4249	0.958	0.852	1.076	0.46714	0.99	0.86	1.141	0.8924	0.927	0.673	1.276	0.64042
REACTOME MOLECULES ASSOCIATED WITH ELASTIC FIBRES	0.958	0.852	1.078	0.47738	0.97	0.843	1.117	0.6748	0.939	0.676	1.306	0.70957
KEGG MEDICUS REFERENCE LPA GNA12 13 RHOA SIGNALING PATHWAY	0.958	0.85	1.08	0.4817	0.907	0.785	1.047	0.1817	1.029	0.741	1.43	0.86423
KEGG SMALL CELL LUNG CANCER	1.044	0.926	1.178	0.48229	1.028	0.887	1.19	0.71639	0.911	0.656	1.267	0.58051
PID EPHA FWDPATHWAY	0.962	0.857	1.08	0.51251	0.965	0.84	1.109	0.61569	1.125	0.813	1.555	0.47712
REACTOME GAP JUNCTION TRAFFICKING AND REGULATION	1.041	0.923	1.175	0.51438	0.992	0.857	1.149	0.91476	1.165	0.833	1.628	0.37198
WP OMEGA 3 OMEGA 6 FATTY ACID SYNTHESIS	1.039	0.924	1.168	0.52045	1.084	0.942	1.249	0.26087	1.2	0.861	1.673	0.28095
REACTOME ANCHORING OF THE BASAL BODY TO THE PLASMA MEMBRANE	0.963	0.858	1.082	0.528	1.024	0.89	1.178	0.738	0.822	0.597	1.131	0.22771
KEGG MEDICUS REFERENCE ACTIVATION OF PRC2.2 BY UBIQUITINATION OF H2AK119 IN GERMLINE GENES	1.037	0.92	1.169	0.55187	1.109	0.96	1.281	0.15934	0.82	0.581	1.156	0.25685
REACTOME VIRAL INFECTION PATHWAYS	1.036	0.921	1.166	0.55266	1.103	0.958	1.271	0.17291	0.84	0.609	1.159	0.28932
REACTOME TCF DEPENDENT SIGNALING IN RESPONSE TO WNT	0.969	0.861	1.091	0.60328	1.025	0.888	1.184	0.73177	0.828	0.594	1.155	0.26671
REACTOME SIGNALING BY RHO GTPASES MIRO GTPASES AND RHOBTB3	1.032	0.913	1.167	0.61453	1.109	0.954	1.289	0.17713	1.011	0.72	1.419	0.94954
WP PARKIN UBIQUITIN PROTEASOMAL SYSTEM PATHWAY	1.03	0.918	1.156	0.61671	1.046	0.909	1.202	0.53094	0.999	0.727	1.373	0.99424
REACTOME MHC CLASS II ANTIGEN PRESENTATION	0.969	0.854	1.098	0.62052	0.98	0.841	1.141	0.79124	0.862	0.609	1.221	0.40343
REACTOME FCGAMMA RECEPTOR FCGR DEPENDENT PHAGOCYTOSIS	1.03	0.915	1.16	0.62415	1.102	0.954	1.272	0.18762	0.941	0.672	1.318	0.72409
Organelle biosynthesis	0.971	0.86	1.095	0.63138	0.943	0.814	1.092	0.43167	0.905	0.649	1.261	0.55422
WMH PRSCS global	1.029	0.903	1.174	0.66609	0.983	0.839	1.152	0.83212	1.06	0.735	1.527	0.75627
REACTOME PROTEIN LOCALIZATION	1.025	0.911	1.154	0.68146	0.995	0.861	1.148	0.94078	1.15	0.826	1.601	0.40681
REACTOME ELASTIC FIBRE FORMATION	0.978	0.87	1.099	0.70693	1	0.868	1.151	0.99593	0.886	0.633	1.239	0.47846
REACTOME ORGANELLE BIOGENESIS AND MAINTENANCE	1.021	0.906	1.151	0.72888	0.994	0.86	1.149	0.93387	0.907	0.655	1.256	0.55681
WP FOCAL ADHESION	1.02	0.905	1.149	0.74655	1.05	0.91	1.212	0.50456	1.019	0.716	1.45	0.91884
REACTOME CHYLOMICRON CLEARANCE	1.022	0.891	1.173	0.75599	1.017	0.861	1.202	0.84237	1.248	0.857	1.818	0.24808
REACTOME RHO GTPASE CYCLE	1.019	0.901	1.152	0.77028	1.076	0.927	1.248	0.33714	0.979	0.69	1.388	0.90412
WP EICOSANOID METABOLISM VIA LIPOOXYGENASES LOX	1.018	0.902	1.148	0.77158	0.983	0.849	1.138	0.81673	1.132	0.803	1.595	0.48032
REACTOME EPIGENETIC REGULATION OF GENE EXPRESSION	1.016	0.902	1.144	0.79458	1.021	0.885	1.18	0.77198	0.907	0.65	1.266	0.56709
WP RESISTIN AS A REGULATOR OF INFLAMMATION	1.016	0.902	1.144	0.80002	1.008	0.872	1.165	0.91701	1.112	0.8	1.544	0.52742
REACTOME SIGNALING BY PDGF	1.015	0.901	1.143	0.80614	0.995	0.861	1.149	0.94197	1.409	1.012	1.963	0.04246
REACTOME G0 AND EARLY G1	0.986	0.87	1.117	0.82311	0.97	0.834	1.128	0.69023	0.843	0.584	1.216	0.36134
PID EPHA2 FWD PATHWAY	0.989	0.876	1.117	0.86475	1.049	0.906	1.216	0.52295	0.903	0.648	1.259	0.54809
WP ESTROGEN RECEPTOR PATHWAY	1.007	0.892	1.136	0.91485	1.061	0.916	1.228	0.43173	1.042	0.741	1.465	0.81417
WP GENES RELATED TO PRIMARY CILIUM DEVELOPMENT BASED ON CRISPR	0.998	0.887	1.124	0.97651	0.975	0.847	1.124	0.7314	0.842	0.597	1.187	0.32527
WP INTEGRIN MEDIATED CELL ADHESION	0.998	0.885	1.126	0.97818	0.973	0.842	1.125	0.71371	1.201	0.849	1.699	0.29997

Supplementary Table 15. *Testing heterogeneity for association with WMH volume between hypertensive sub-groups in the UK Biobank*

PRS	Heterogeneity P	BETA_HTA	CI_LOW_HTA	CI_HIGH_HTA	P_HTA	BETA_noHTA	CI_LOW_noHTA	CI_HIGH_noHTA	P_noHTA
WP OMEGA 3 OMEGA 6 FATTY ACID SYNTHESIS	0.002	0.052	0.034	0.071	3.53E-08	-0.001	-0.028	0.027	9.52E-01
REACTOME CHYLOMICRON CLEARANCE	0.022	0.038	0.019	0.056	5.90E-05	-0.001	-0.029	0.027	9.26E-01
Carbohydrate metabolism	0.024	0.061	0.043	0.08	1.20E-10	0.023	-0.005	0.05	1.04E-01
Protein Localization and Fatty acid metabolism	0.028	0.054	0.036	0.073	1.22E-08	0.017	-0.01	0.044	2.29E-01
REACTOME FCGAMMA RECEPTOR FCGR DEPENDENT PHAGOCYTOSIS	0.032	0.049	0.031	0.068	1.79E-07	0.012	-0.016	0.04	4.06E-01
REACTOME PARASITE INFECTION	0.032	0.041	0.023	0.06	1.06E-05	0.004	-0.024	0.032	7.87E-01
REACTOME PROTEIN LOCALIZATION	0.033	0.048	0.029	0.067	5.31E-07	0.012	-0.015	0.039	3.83E-01
PID EPHA2 FWD PATHWAY	0.068	0.03	0.011	0.048	1.50E-03	-0.001	-0.029	0.027	9.54E-01
REACTOME HIV INFECTION	0.077	0.056	0.038	0.074	2.43E-09	0.026	-0.002	0.054	7.25E-02
KEGG MEDICUS REFERENCE ACTIVATION OF PRC2.2 BY UBIQUITINATION OF H2AK119 IN GERMLINE GENES	0.093	0.039	0.021	0.058	2.66E-05	0.01	-0.018	0.038	4.70E-01
KEGG PEROXISOME	0.096	0.036	0.017	0.055	1.71E-04	0.008	-0.02	0.035	5.83E-01
REACTOME CALCITONIN LIKE LIGAND RECEPTORS	0.112	0.029	0.01	0.047	2.51E-03	0.002	-0.026	0.03	8.88E-01
REACTOME SIGNALING BY VEGF	0.112	0.051	0.032	0.069	6.95E-08	0.024	-0.004	0.052	9.49E-02
PID EPHA FWDPATHWAY	0.126	0.037	0.018	0.055	1.09E-04	0.011	-0.017	0.039	4.33E-01
WP PARKIN UBIQUITIN PROTEASOMAL SYSTEM PATHWAY	0.131	0.029	0.01	0.047	2.47E-03	0.004	-0.024	0.031	7.86E-01
KEGG AMINO SUGAR AND NUCLEOTIDE SUGAR METABOLISM	0.131	0.043	0.024	0.061	8.20E-06	0.018	-0.01	0.045	2.07E-01
REACTOME TCF DEPENDENT SIGNALING IN RESPONSE TO WNT	0.147	0.033	0.014	0.051	5.76E-04	0.057	0.03	0.084	4.70E-05
REACTOME ALPHA LINOLENIC OMEGA3 AND LINOLEIC OMEGA6 ACID METABOLISM	0.154	0.033	0.014	0.052	6.35E-04	0.009	-0.018	0.036	5.07E-01
WP FOCAL ADHESION	0.164	0.054	0.036	0.073	1.10E-08	0.03	0.002	0.058	3.42E-02
WMH PRS global	0.176	0.208	0.189	0.226	3.46E-109	0.185	0.157	0.213	4.14E-38
Metabolism of nucleotide sugars and glycolysis	0.187	0.049	0.03	0.067	2.92E-07	0.026	-0.002	0.055	6.74E-02
REACTOME VIRAL INFECTION PATHWAYS	0.194	0.065	0.047	0.084	7.62E-12	0.042	0.014	0.071	3.12E-03
KEGG MEDICUS REFERENCE MICROTUBULE DEPOLYMERIZATION	0.195	0.025	0.006	0.043	9.09E-03	0.047	0.019	0.075	1.13E-03
REACTOME MOLECULES ASSOCIATED WITH ELASTIC FIBRES	0.227	0.059	0.041	0.077	2.93E-10	0.039	0.011	0.066	6.25E-03
REACTOME ELASTIC FIBRE FORMATION	0.239	0.058	0.039	0.076	7.22E-10	0.038	0.01	0.066	8.01E-03
WP SPHINGOLIPID METABOLISM OVERVIEW	0.251	0.042	0.023	0.06	1.07E-05	0.061	0.033	0.088	1.90E-05
REACTOME GAP JUNCTION TRAFFICKING AND REGULATION	0.259	0.04	0.021	0.059	2.62E-05	0.021	-0.007	0.048	1.43E-01
KEGG SMALL CELL LUNG CANCER	0.271	0.03	0.012	0.049	1.24E-03	0.049	0.021	0.077	5.42E-04
WP INTEGRIN MEDIATED CELL ADHESION	0.289	0.049	0.03	0.067	2.28E-07	0.031	0.003	0.059	3.15E-02
REACTOME VITAMIN B5 PANTOTHENATE METABOLISM	0.313	0.036	0.017	0.055	1.59E-04	0.053	0.025	0.08	1.62E-04
Immune response to viral infections	0.317	0.076	0.057	0.094	1.19E-15	0.059	0.031	0.087	4.43E-05
REACTOME INFECTIOUS DISEASE	0.317	0.076	0.057	0.094	1.19E-15	0.059	0.031	0.087	4.43E-05
REACTOME SIGNALING BY RHO GTPASES MIRO GTPASES AND RHOBTB3	0.325	0.065	0.047	0.084	5.38E-12	0.048	0.02	0.076	8.77E-04
REACTOME DISEASES ASSOCIATED WITH GLYCOSAMINOGLYCAN METABOLISM	0.354	0.043	0.025	0.062	4.90E-06	0.027	-0.001	0.055	5.73E-02
WP CONGENITAL GENERALIZED LIPODYSTROPHY	0.354	0.043	0.025	0.062	5.39E-06	0.027	-0.001	0.055	5.47E-02
WP HEDGEHOG SIGNALING PATHWAY WP4249	0.354	0.021	0.003	0.04	2.36E-02	0.037	0.009	0.065	8.71E-03
KEGG MEDICUS REFERENCE LPA GNA12 13 RHOA SIGNALING PATHWAY	0.373	0.037	0.019	0.056	9.00E-05	0.052	0.024	0.079	2.51E-04
WP SPHINGOLIPID METABOLISM INTEGRATED PATHWAY	0.385	0.042	0.023	0.061	1.01E-05	0.057	0.029	0.085	5.38E-05
KEGG MEDICUS REFERENCE REGULATION OF GF RTK RAS ERK SIGNALING RAS UBIQUITINATION BY CUL3 COMPLEX	0.398	0.026	0.007	0.044	6.49E-03	0.012	-0.016	0.039	4.10E-01
REACTOME MHC CLASS II ANTIGEN PRESENTATION	0.444	0.037	0.018	0.055	1.13E-04	0.024	-0.004	0.052	9.65E-02
BIOCARTA NPP1 PATHWAY	0.469	0.031	0.012	0.049	1.22E-03	0.043	0.016	0.07	2.11E-03
PID TOLL ENDOGENOUS PATHWAY	0.553	0.033	0.015	0.052	4.67E-04	0.023	-0.005	0.05	1.05E-01
REACTOME ORGANELLE BIOGENESIS AND MAINTENANCE	0.562	0.048	0.03	0.067	4.02E-07	0.058	0.03	0.086	5.04E-05
REACTOME SIGNALING BY PDGF	0.562	0.043	0.025	0.062	5.35E-06	0.033	0.006	0.061	1.84E-02
WP RESISTIN AS A REGULATOR OF INFLAMMATION	0.562	0.035	0.017	0.054	1.94E-04	0.025	-0.003	0.053	7.95E-02
Signal Transduction	0.596	0.068	0.049	0.086	5.64E-13	0.059	0.031	0.087	3.87E-05
PID AVB3 OPN PATHWAY	0.602	0.027	0.009	0.046	4.07E-03	0.036	0.008	0.064	1.20E-02
WP EICOSANOID METABOLISM VIA LIPOOXYGENASES LOX	0.602	0.03	0.012	0.049	1.49E-03	0.039	0.011	0.067	5.88E-03
REACTOME SIGNALING BY RECEPTOR TYROSINE KINASES	0.638	0.065	0.047	0.083	4.14E-12	0.057	0.03	0.085	5.51E-05
WP DISORDERS OF GALACTOSE METABOLISM	0.685	0.032	0.013	0.051	7.36E-04	0.025	-0.003	0.053	7.76E-02
WP GENES RELATED TO PRIMARY CILIUM DEVELOPMENT BASED ON CRISPR	0.722	0.052	0.033	0.071	5.28E-08	0.046	0.019	0.073	8.90E-04
Organelle biosynthesis	0.724	0.061	0.042	0.079	1.26E-10	0.067	0.04	0.095	1.96E-06
REACTOME RHO GTPASE CYCLE	0.724	0.059	0.04	0.077	5.11E-10	0.053	0.025	0.081	2.21E-04
REACTOME EPIGENETIC REGULATION OF GENE EXPRESSION	0.728	0.043	0.025	0.062	5.01E-06	0.037	0.01	0.065	8.29E-03
REACTOME MITOCHONDRIAL TRANSLATION	0.763	0.033	0.014	0.051	6.21E-04	0.038	0.01	0.065	7.21E-03
REACTOME G0 AND EARLY G1	0.768	0.035	0.016	0.053	2.30E-04	0.04	0.012	0.068	5.52E-03
WP ESTROGEN RECEPTOR PATHWAY	0.768	0.028	0.009	0.046	3.76E-03	0.023	-0.004	0.051	9.32E-02
REACTOME ANCHORING OF THE BASAL BODY TO THE PLASMA MEMBRANE	0.772	0.044	0.025	0.063	3.69E-06	0.049	0.021	0.077	5.88E-04
REACTOME EPH EPHRIN SIGNALING	0.856	0.031	0.012	0.049	1.10E-03	0.028	0	0.055	4.96E-02
REACTOME PRE NOTCH PROCESSING IN GOLGI	0.860	0.025	0.007	0.043	7.50E-03	0.022	-0.007	0.05	1.36E-01
REACTOME TRANSCRIPTION OF E2F TARGETS UNDER NEGATIVE CONTROL BY DREAM COMPLEX	0.863	0.033	0.014	0.051	4.99E-04	0.03	0.002	0.059	3.72E-02
REACTOME EPHA MEDIATED GROWTH CONE COLLAPSE	0.908	0.026	0.008	0.045	5.85E-03	0.024	-0.004	0.052	8.91E-02