UNIVERSITÉ PARIS SCIENCES & LETTRES

Thèse de doctorat *de l'Université PSL*

Préparée à l'Université Paris Dauphine-PSL

Reaching the Hard-to-Reach: Essays on Integration, Juvenile Justice, and Access to Care

Soutenue par

Alexandre Touw

Le 12 octobre 2026

École doctorale n°543

SDOSE

Spécialité

Sciences Économiques

Composition du jury :

Roberto GALBIATI Directeur de recherche CNRS Sciences Po	*Président du jury*
Liam WREN-LEWIS Directeur de recherche INRAE Paris School of Economics	*Rapporteur*
Mette FOGED Associate Professor & Senior Researcher University of Copenhagen (UCPH) Rockwool Foundation	*Examinatrice*
Arnaud PHILIPPE Associate Professor University of Bristol	*Examinateur*
Flore GUBERT Directrice de recherche IRD Université Paris Dauphine-PSL	*Directrice de thèse*
Élise HUILLERY Professeure des Universités École Normale Supérieure-PSL Université Paris Dauphine-PSL	*Directrice de thèse*

Dauphine UNIVERSITÉ PARIS SCIENCES & LETTRES

L'Université Paris Dauphine - PSL n'entend donner aucune approbation ni improbation aux opinions émises dans les thèses ; ces opinions doivent être considérées comme propres à leurs auteurs.

Reaching the Hard-to-Reach: Essays on Integration, Juvenile Justice, and Access to Care

Alexandre Touw

À Martine et Dadda Amghar,

Remerciements

Au cours de ces cinq dernières années, on m'a souvent demandé si je ne me sentais pas un peu isolé dans ce travail de thèse et quelle mouche avait bien pu me piquer pour que je décide de me lancer dans un bourbier pareil. Si je ne doute pas que cette expérience puisse être celle de certains doctorants, je n'ai jamais vécu cette thèse comme un travail solitaire, simplement parce que cela ne l'a (quasiment) jamais été pour moi. Et bien que j'ai tout de même sûrement du perdre le (peu) de cheveux qu'il me restait en cours de route, j'ai eu la chance d'être constamment entouré, dans mon travail comme en dehors, par des personnes qui m'ont soutenu et rassuré. Si je garde de cette thèse un souvenir aussi heureux, c'est en immense partie grâce à elles. Je suis heureux de pouvoir commencer ces pages en prenant le temps de les remercier, de manière volontairement extensive, comme elles le méritent[1].

Mes premiers remerciements vont naturellement à mes deux directrices de thèse, Élise et Flore. J'ai énormément de mal à imaginer à quoi auraient ressemblé ces (presque) sept dernières années sans vous deux.

Élise, quand je t'ai rencontrée, je n'étais encore qu'un ingénieur un peu perdu qui avait candidaté plus ou moins par hasard, et sans vraiment penser être retenu, à un poste d'assistant de recherche au LIEPP. Tu m'as fait confiance une première fois (je me demande d'ailleurs toujours pourquoi) et ce premier poste a changé la vision que j'avais sur le métier de chercheur. Nous avons très rapidement discuté d'un doctorat et tu m'as fait confiance une seconde fois en me présentant à Flore et Jean-Noël pour travailler sur un projet qui allait devenir, quelques années et quelques aventures plus tard, le premier chapitre de cette thèse. Si, depuis, nos relations ont dépassé le simple cadre de travail, allant de débats (parfois trop ?) animés à la Kfet de la rue d'Enghien aux week-ends confessions à la campagne en passant par des blind tests drag, tu as toujours pris soin de dégager du temps quand j'en avais besoin. J'espère que tu ne m'en voudras pas, certaines figures ont mystérieusement réapparu dans le corps des papiers ...

Flore, notre relation a commencé avec OEPRE et je garde un souvenir attendri de ma rencontre avec Jean-Noël et toi au Mauri7 ... sans Jean-Noël (promis, ça sera la seule blague sur ton légendaire enthousiasme matinal JN). Si les thématiques migratoires ont toujours été un de mes centres d'intérêt de recherche, ce projet et nos discussions m'auront conduit à m'y tourner plus franchement, au point qu'elles occupent aujourd'hui la majorité de mes projets de recherche, tous à tes côtés pour l'instant. Tu as énormément compté dans cette évolution personnelle, et je t'en remercie. J'ai aussi eu la chance de partager ton bureau pendant les deux dernières années de cette thèse et, de manière contre-intuitive pour un doctorant, être enfermé à un mètre de sa directrice de thèse n'est finalement pas si terrible que ça. Merci pour les

[1] While the unusual length of this acknowledgements section genuinely reflects how much I owe to all those mentioned in it, the fact that the table of contents only comes afterwards may prove somewhat inconvenient for readers (members of the jury, for instance) particularly eager for some scientific reading. They may therefore wish to skip directly to page xxiii, where they will find the table of contents.

discussions famille, les concerts, et les échanges nocturnes d'insomniaques, tu as toujours su me rassurer et me rappeler ce qui importait vraiment.

À toutes les deux, je veux surtout vous dire un immense merci. Vous avez toujours trouvé le temps lorsque j'en avais besoin, et su me soulager lorsque j'en avais trop sur les épaules, que cela soit pour le travail comme pour les moments plus difficiles. Co-écrire avec vous m'a permis de structurer la manière dont je me pose des questions et dont je les aborde, j'ai appris énormément à vos côtés et ne pouvais rêver meilleur binôme d'encadrement. J'espère que la fin de cette thèse appellera d'autres projets (je crois que JN a quelque chose sur le feu), et que j'aurai l'occasion de continuer à apprendre à vos côtés.

I am deeply grateful to the members of my jury: Mette Foged, Roberto Galbiati, Arnaud Philippe, and Liam Wren-Lewis. Mette, although we never had the opportunity to meet in person (yet), I am very happy to have the chance to discuss this work with you. We share many research interests, and I feel very fortunate to benefit from your insight. I am all the more grateful that you agreed to take part despite the timing, and to spend part of your holidays discussing this thesis. Arnaud, having you on the committee felt like the most natural thing. You have been there since the very beginning, even before we had met in person and before I actually started my PhD, and I am particularly happy that you will be there for this important step.

Roberto and Liam, you have already read these chapters for my pre-defence, and a good part of the work that followed was guided by your comments. I am therefore doubly grateful for this additional round of reading, and the time and attention I am asking of you again.

Je suis également reconnaissant envers l'Université Paris Dauphine et l'École Doctorale (et ses différentes directions, Élise Huillery, Farid Toubal et Anne Épaulard ainsi qu'Hubert Radenac pour sa flexibilité dans le calendrier de cette soutenance), le programme gradué (Eve Caroli et David Ettinger) et la direction de DIAL (Anne-Sophie et Jalil & Kenneth) pour m'avoir permis d'effectuer ma thèse dans de bonnes conditions de travail.

Au LEDa spécifiquement, je souhaite remercier Philippe de Vreyer, ainsi que Vanina et Cécile. Cécile, ton soutien administratif dans la réalisation de ces projets a été déterminant plus d'une fois, et dans un monde parfait tu mériterais plus que ces quelques lignes pour chanter tes louanges. Alors un grand merci d'avoir fait en sorte que j'appréhende toujours le suivi financier d'un projet sereinement, merci pour nos échanges toujours agréables, nos partages de photos de chat et de vacances. Je passe la main à Hala pour le reste !

Je suis également très reconnaissant envers Marie-Pierre Dargnies, José Trashorras, Marie Bessec et Joachim Jarreau de m'avoir tous les quatre fait confiance pendant ces quelques années pour l'enseignement des statistiques. Votre disponibilité, vos conseils et matériaux ont grandement contribué à ce que ce soit une expérience agréable.

Je tiens également à remercier Alice Mesnard et Arnaud Philippe (une deuxième fois) pour ces rendez-vous annuels du comité de thèse qui m'ont toujours obligé à prendre un peu de recul et à remettre de l'ordre dans mes idées et dans cette thèse. Mais c'est également le bon moment pour moi de vous remercier pour tous nos échanges plus informels tout au long de l'année. Alice,

entre DIAL, le Bal Blomet et les recherches de financements, tu n'as jamais été avare de conseils et d'idées, et je t'en remercie. Arnaud, nos cafés au ministère ont été beaucoup trop rares, mais toujours précieux pour m'aider à naviguer dans les subtilités, parfois assez mystérieuses, d'une administration centrale. Vos conseils à tous les deux m'ont été précieux, et souvent aussi beaucoup rassuré.

Ce manuscrit est composé aux deux tiers de travaux co-écrits. Si je ne me suis jamais vraiment senti seul ou dépassé par tout ce qu'il y avait à faire, c'est en grande partie parce que j'ai eu la chance de travailler avec des co-auteurs investis et passionnants.

Jean-Noël, ta contribution à cette thèse dépasse largement notre co-écriture. Tes conseils m'ont souvent permis de reprendre du recul sur ce travail, et de revenir sur ce qui était vraiment important. On ne partage pas seulement la même directrice de thèse, mais aussi suffisamment d'intérêts de recherche pour que le titre de cette thèse ait, à un moment, dangereusement ressemblé à celui de ton HDR. Je mesure la chance que j'ai eue de pouvoir compter sur toi tout au long de cette thèse.

Jérôme, travailler avec toi m'aura au moins appris qu'on peut être d'une efficacité redoutable, ne jamais vraiment compter son temps pour faire avancer un projet, sans pour autant oublier qu'il existe une vie autour ... ni renoncer à aller boire un verre une fois le travail terminé. Merci pour tous tes conseils sur la recherche, la rédaction, le milieu et ses quelques règles (comme celle de toujours mettre ses vecteurs en gras). Enfin merci grandement pour ce petit message qui m'a permis, un jour, de faire à un certain A.C une annonce qu'il n'est pas près d'oublier de sitôt. Marie-Caroline, je suis très heureux d'avoir eu la chance de travailler avec toi. Ton regard de psychologue clinicienne et ta connaissance extrêmement fine de la santé mentale, de l'hôpital et des parcours des personnes réfugiées ont apporté au projet une dimension que nous n'aurions évidemment jamais pu avoir seuls. J'apprends encore beaucoup à chacune de nos discussions, et je suis très reconnaissant de pouvoir faire ce projet à tes côtés. Hala, je suis particulièrement heureux de savoir que l'aventure ne fait que commencer pour toi. Tu fais partie de ces personnes qui prennent chaque nouveau problème à bras-le-corps, avec un sens de l'organisation et une rigueur qui continuent assez régulièrement de me fasciner. Je suis donc ravi que la suite se passe à Dauphine avec Flore, et surtout que cette année passée avec nous ait pu te convaincre de te tourner vers la thèse (on n'y est pas allé de main morte non plus). Cela me donnera une excellente excuse pour continuer à passer à DIAL, et peut-être enfin assister à l'un de ces fameux ateliers d'origami que je rate constamment.

Et puis, il y a les projets qui sont dans cette thèse, et ceux qui restent encore un peu dans l'ombre (parfois depuis un certain temps).

Éléonore, l'heureux hasard d'un appel à projet aura suffi à nous mettre sur le même chemin. Ta manière d'appréhender les questions de pauvreté et de vulnérabilité fait du bien, et j'ai hâte que notre projet commence vraiment. J'ai d'autant plus hâte parce que je lui dois déjà quelque chose d'important : il aura aussi permis de remettre Laudine sur ma route. Laudine, on s'était un peu perdus de vue depuis notre rencontre au LIEPP, et je suis très heureux que cette fin de thèse nous ait donné l'occasion de nous retrouver, qui plus est autour de ce projet "on ne peut

plus LIEPP" sur l'aide sociale à l'enfance qui me tient particulièrement à cœur. J'ai vraiment hâte de travailler avec vous deux.

Merci aussi à toute l'équipe d'AlloSchool : Alice, Flore, Séverine, Cynthia, Noé, Coralie, Franck et Marie-Line. Voir l'équipe s'agrandir au fil du temps nous conforte au moins dans l'idée que ce projet mérite qu'on s'accroche.

Enfin, je remercie Carlo et Pauline qui êtes, avec Élise, techniquement mes tout premiers vrais co-auteurs. Il est d'ailleurs peut-être temps, après toutes ces années, d'envisager enfin de fermer la porte de ce projet (je prends le risque de l'écrire ici pour forcer le destin et que cela se réalise enfin).

Ces travaux n'auraient pas été possibles sans l'accès aux données administratives de plusieurs ministères et, évidemment, sans les personnels qui facilitent ces démarches.

Au ministère de l'Éducation nationale, je tiens particulièrement à remercier le bureau de l'appui à l'évaluation des politiques publiques et de soutien à la recherche ainsi que le programme IDEE. Axelle, tu m'as fait confiance également une première fois avec Clémence pour venir travailler quelque temps au bureau B4 et cette expérience a forgé ma gratitude pour la statistique publique et je suis admiratif de tout le travail que tu as accompli avec ce nouveau bureau. Tu m'as toujours soutenu, parfois dans des contextes plus délicats, et je t'en remercie sincèrement. Alexis, je te remercie pour ton infinie patience, ta flexibilité et ta rapidité. J'ai essayé de ne pas en abuser ... et je pense pouvoir affirmer avoir échoué. Tu peux donc considérer ces remerciements comme des excuses pour toute la gestion administrative, souvent laborieuse, de nos conventions. Sylvia, pour ta bonne humeur communicative et ton aide précieuse dans les mises à disposition de données. Puisque je sais que vous faites votre maximum pour le développer, je tiens à inscrire sur papier que je me sens très chanceux d'avoir pu profiter de la DEPP@thèque au cours de cette dernière année de thèse, cela m'a permis de vivre ce sprint final de manière beaucoup plus sereine. Ce type d'outils mériterait d'être pris comme exemple de ce que la statistique publique peut faire pour rendre accessible les données aux chercheurs de manière plus fluide mais surtout plus égalitaire, et je suis admiratif de votre engagement sur ce sujet et reconnaissant de votre considération de la situation des jeunes chercheurs. Daniel, tu fais la transition avec le programme IDEE et je te remercie pour le temps passé à préparer nos données. Quentin, ton aide méthodo et tes suggestions de chercheurs en psycho auront permis de créer une équipe interdisciplinaire pour un projet entier, je t'en suis extrêmement redevable. Alexandre, nos échanges pendant nos passages en salle chercheurs et ton appui dans le développement d'outils pour IDEE m'ont beaucoup aidé dans les prémices de certains projets. Je t'en suis reconnaissant et suis heureux de te savoir, sans surprise, en thèse. Enfin, un grand merci à Clémence. Mon passage en tant qu'assistant de recherche à tes côtés a profondément marqué ma manière de travailler au quotidien et je peux enfin profiter de ces remerciements pour rétablir la maternité de ma rigidité sur l'organisation des codes, et notamment sur l'utilisation de R (et surtout RMarkdown). Merci également pour tes mises en relation dans les contextes de nos projets et tes propositions d'interventions toujours bienvenues. Nos déjeuners me manquent et j'espère qu'on aura le temps de se voir, malgré de petits (et bruyants) événements à venir.

Au ministère de la Justice (PJJ et SSER), je tiens à remercier Aurélie Fillod-Chabaud qui

a permis que ce projet sur les CEFs mette un pied dans la porte, ainsi que Sylvain Renault et Isabelle Thing-Leoh pour leur expertise sur les données. L'accès aux données n'aurait jamais pu voir le jour sans l'accompagnement d'Alice Simon, Pascal Chevalier et Malik Koubi, je les en remercie sincèrement. J'ai également eu la chance de pouvoir travailler dans les locaux du SSER, et je profite de ces lignes pour remercier Luc, Élise et Gabriela pour l'accueil dans le bureau (des potins). Vous avez rendu mes (malheureusement trop rares) excursions au Millénaire aussi joyeuses qu'animées, je vous en remercie. Merci à Zakia et Yara pour leur accueil chaleureux (et les sucreries) et à Olivier pour son aide logistique et sa bonne humeur. Maintenant que mon agenda semble soulagé d'un poids, j'espère bien revenir plus souvent ... surtout les mardis !

Au ministère de l'Intérieur (DSED-DGEF), je remercie chaleureusement Samuel Ettouati pour tout le travail de facilitation d'accès aux données pour les chercheurs, et de coordination de projets.

En complément de ces données administratives, cette thèse repose également beaucoup sur des données collectées sur le terrain. Dans le cadre d'OEPRE, je remercie donc tous les responsables de CASNAV ayant pris sur leur temps pour nous aider dans la phase d'échantillonage et de prise de contact avec les écoles, ainsi que Thibaut Mainville, Émeline Bon, Lucie Roy, Stéphanie Véloso et Clément d'Inguimbert à la DGESCO et Philippe Arrondeau, Valérie Gallat, Virginie Jourdan et Eugénie Marie pour le Ministère de l'Intérieur.

Pour l'implémentation de la RCT dans les locaux de France Terre d'Asile-AGIR à Paris, nous avons pu compter sur le soutien de Delphine Rouilleault, Véronique Otchoumou, Victoire Larzillière et tous les travailleurs sociaux. À Coallia dans l'Essonne, l'aide d'Edouard Osternaud, Jacques Tome Bodiong, Aissetou Conteh et toutes les travailleuses sociales a été déterminante. Nous les remercions pour leur temps et les locaux qu'ils ont bien voulu nous dédier.

J'ai eu la chance de passer presque six ans dans les locaux de la rue d'Enghien de l'équipe DIAL du LEDa, et je tiens chaleureusement à remercier toutes les personnes qui ont contribué à faire de ce marathon une promenade de santé au quotidien. En tout premier lieu Danielle, la gardienne des lieux et du café, et sauveuse officielle des plantes du labo. Je ferai en sorte d'avoir toujours une photo de Maurice à te montrer quand je repasserai. Merci à Anne et Loïc pour l'accompagnement administratif pas toujours évident à gérer.

S'il est vrai que la géographie du lieu fait beaucoup dans la facilitation des interactions sociales, le reste découle largement de la personnalité des chercheurs. Je tiens donc à remercier (en partant du bureau du fond en date d'août 2026) Sandrine pour sa bonhommie et ses compétences de chasseuse de pigeons à ses heures perdues, Anne-Sophie pour nos discussions musicales, Kenneth pour tes encouragements et ton soutien dans nos épreuves avec nos ministères respectifs, Yannick pour ton sens de l'humour et tes fun facts animant nos déjeuners à la Kfet, Camille pour être la seule capable de partager le bureau de JN, JN (encore) pour les pauses clopes réconfortantes, Marin pour la bonne humeur constante que tu propages si facilement, Anne-Laure pour nos discussions politiques, la rédaction d'un tiers de cette thèse dans un cadre idéal, et surtout, surtout ... pour avoir retiré la photo traumatisante de B.V de ce groupe WhatsApp, Véronique pour m'avoir convaincu de marcher en chaussettes au labo et nos discussions Claude,

Marion pour ton oreille attentive aux conditions des doctorants et ta joie de vivre et enfin Thomas, pour ta sollicitude, et ton soutien émotionnel à chaque étape difficile de ces dernières années.

J'ai été entouré, année après année, par des doctorants devenus de vrais compagnons de route. Je pense d'abord à ceux qui ont déjà vogué vers de nouvelles aventures et tiens donc à remercier Yvan (je comprends maintenant ton désemparement administratif à l'approche de la deadline du rendu mais on va la décrocher cette thèse de gestion), Antoine (ça sera peut-être plus simple de se lancer dans un projet avec Antoine maintenant), Mary (pour m'avoir montré ce que ténacité veut dire), Paul Br. (pour ces discussions matching), Mathilde S. (pour ton soutien pendant les calls for proposals, les petites bières bruxelloises et surtout les photos de chat) et Paul Bl. (sache que ta passion pour les petites collations sucrées aura finalement été transmise et j'écris ces lignes avec un carré de choco fleur de sel en bouche, merci pour avoir rendu ces derniers mois ponctués de pauses café, de qualité c'est important, plus agréables) les Tic & Tac du bureau AER/Chocolat, Fatou (für deine ansteckend gute Laune, deinen unermüdlichen Einsatz für die Promovierenden und für deine immer skurrileren Träume. PS: Ich glaube, inzwischen träume ich ab und zu ... ich bin also doch kein Soziopath), Mathilde L. (pour les fun facts surprenants de scandales criminels français, les galères de correction de copies, et les photos de Crabi), et Jeanne (pour avoir été l'élément moteur de la rue d'Enghien et du bureau du fond pendant tant d'années, un soutien sans faille dans les moments compliqués ... et aussi, quand même, pour ces quelques pastis).

J'ai une pensée émue pour tous les doctorants avec lesquels j'ai partagé ces dernières années de thèse et qui sont encore dedans. Komlavi (pour toutes ces luttes administratives et cette petite sieste, un jour, sur mon canapé), Adrien (on se sera rencontré tardivement mais je suis content d'avoir pu partager ces petits moments avec toi et suis admiratif de ton parcours), Tiavina (ça aura pris le temps depuis les bancs d'Assas mais je crois quand même que ça valait un peu le coup !), Malte (pour ta capacité à organiser des séminaires comme prétexte pour aller boire une bière derrière), Lucia (pour tous ces petits moments de "débrief" des cours de L3 et ces discussions, disons, ... décomplexées) et Mélodie (pour avoir réussi à ne plus jamais remettre les pieds rue d'Enghien malgré les projets de resto kurdes et tes recommandations d'influenceurs peu recommandables), Angela (pour cette découverte inopinée de ta passion cachée à la Sirène), Paul-Émile (pour squatter les biens publics #rendsletemplatedethèse, les photos - risquées - à l'autre bout du monde, le soutien moral dans les corrections de copies, me montrer ce que c'est de la vraie économie, les procédures administratives ubuesques, mais surtout pour le rhum arrangé. Tu vois, je t'avais dit que je les retrouverai ces remerciements), Mathilde B. (pour ton investissement à la communauté du labo, ton rire communicatif, partager ces cours avec toi a été un vrai plaisir), Mathilde (je crois que c'est la dernière ?) P. (pour ta détermination, nos recos de bouquins et des raclettes qui dérapent), Antoine C. - c'est pas encore toi Castet, tu vas devoir patienter encore un peu - (pour m'avoir montré qu'il était possible d'offrir des boucles d'oreilles guillotines à sa partenaire), Joséphine (grâce à toi, je me mets aux jeux à gratter), João (soutien à toi, bring back the cat !), Marta (pour tes histoires de terrain), Hanna (je suivrai ton aventure au moins aussi importante que la thèse: l'arrêt de la clope), Laïla (pour ton rire communicatif,

nos galères de gestion de terrain, de données; il était vraiment temps que tu rentres en thèse !), Hugo (pour toutes ces idées de projets sur des populations vulnérables, pour arriver à me faire douter des procédures administratives que je pensais maîtriser, mais surtout pour Farsou) et Lou (la queen des confs, la digne héritière de Jeannette, pour la bonne humeur que tu insuffles au bureau et ton soutien indéfectible. La bise à Ponyo), et Florence (un immense merci d'avoir toujours été là dans les moments difficiles, pour les échanges de photos de safari, et pour m'avoir montré que, au moins dans certains cas, il est bon de garder la foi. Je la garde aussi pour que tu réussisses enfin à venir voir Maurice). J'ai eu beaucoup de chance de croiser votre route et je vous envoie tout mon soutien pour la fin de cette étape.

Un grand merci de last minute aux postdocs méga stylés du LEDa, Pol et Romane, pour le soutien jusqu'au bout de cette épreuve, et ce, malgré notre rencontre tardive. Je suis heureux de savoir que la lutte continue avec vous à coups de godets, discussions barricades et bitchage sur la situation de l'ESR.

Antoine C. (c'est bon, c'est bien pour toi maintenant) le patron, je ne m'imaginais pas que ce crochet par le LIEPP aurait tant d'impact sur ma trajectoire, et je n'aurais pas pu imaginer meilleur compagnon de route. Nos petits verres réguliers à l'Engrenage m'auront toujours permis d'y voir un peu plus clair sur ce qui m'attendait, et ton soutien à chaque étape, du projet de thèse aux candidatures de postdoc, aura été déterminant. Merci d'essayer d'alléger un peu mon syndrome de l'imposteur, je suis heureux de t'avoir à mes côtés et j'ai hâte de venir te voir avec Camille et JB (le king) à Rouen.

Cette section a commencé comme une tentative de liste volontairement exhaustive de gens à remercier, et je me rends compte, six pages plus loin, qu'elle ne respecte pas vraiment les longueurs traditionnelles de remerciements. Jean-Noël, tu m'as dit un jour que si je devais remercier tout le monde, les remerciements feraient quinze pages et que ça prendrait trois jours . . . tu avais encore raison. Il est cependant trop tard pour revenir en arrière et je vais suivre ton conseil : « C'est ta thèse, tu fais ce que tu veux ».

Le temps passe, la prépa date d'il y a maintenant bien plus de dix ans, et certains génies restent. Un grand merci à Aden pour son intérêt pour ma recherche, son développement capillaire inversement proportionnel au mien, et ses tentatives de me vendre Saint-Étienne comme ville d'avenir, à Slim pour nos discussions musicales, ses histoires de globe-trotter, et ses visites guidées de Rouen, à Matteo (et Julia !) pour les verres inopinés et les siestes devant Wagner, à Rémi pour les verres au Hurling, les propositions de fromage jamais récupérés, et les photos surprises avec Margueb, à Pierre (ou Pentz je sais plus) pour ses mix qui ont accompagné mes soirées depuis le confinement, à Anouchka pour ses conseils immobiliers et ne pas me laisser seul avec un nom marrant, à Nao (et Pascal) pour avoir supporté Aden toutes ces années à ma place, à DLPDV pour son accueil à bras ouverts pendant mes (courtes) aventures lyonnaises et à Lucile pour ses annonces surprises en route pour un week-end perchois mouvementé (bisous Charlie !).

Après la prépa, il y a l'école. J'ai eu la chance d'apprendre auprès des plus grands intellectuels

et de toucher à différents courants de pensée. The F. d'abord, je suis désolé, j'ai fait mes choix et j'ai du couper un peu dans les histoires, je n'avais pas la place pour un chapitre de plus.
Thibaud, il aura suffi d'une référence Kaamelott le deuxième jour pour que tout s'enclenche; les tentatives d'homicides sur personnes asthmatiques, les idées débiles avec Perdig, les décisions de traverser l'Atlantique, les randonnées matinales et très enneigées au parc Lafontaine (un peu coûteuses pour moi mais qui t'auront permis de faire de belles rencontres), les Burger ... Quiz, et autres aventures. Nos retrouvailles sont toujours pour moi un moment précieux et, si l'Aquitaine paraissait bien loin pendant cette thèse, elle me parait étrangement un peu plus accessible maintenant qu'elle est finie.
Alexandre, notre rencontre m'a fait l'effet d'une claque et suis immensément flatté par l'honneur que vous me faites avec Colette. Il me reste néanmoins une vieille dette envers toi : un dernier TP d'élasticité, dont le sujet se trouve ici, vois ça comme mon premier cadeau de mariage.
Pierre, j'avais évidemment écrit quelques lignes pour toi. Malheureusement, il semblerait que notre grande proximité ces derniers temps m'ait conduit à attraper ta maladie et que j'ai, de manière incompréhensible, perdu ces remerciements[2].

Hugo, dès nos vingt ans tu rassemblais les nôtres,
Au BDA déjà tu guidais tous les autres.
Tu tiens notre tribu d'un invisible fil,
Sans toi, chaque soirée irait droit au péril.
Pour Corinne, un beau jour, la Liberté prit vie,
Nos chants et nos portraits célébraient notre amie.
Du clavier à la voix, tu fais vibrer les corps,
Tu veux que mon archet se joigne à tes accords.
Quand l'orchestre exigeait mon grand calme au pupitre,
Tu glissais des clichés qu'on tait dans une épître.
Puis Jordan est venu, confirmant ton bon goût,
Boulogne, hélas, vous tient bien trop loin, tout au bout.
Besançon t'a formé, Paris veut te garder,
Pour jouer, je promets de ne plus tant tarder.
Mes vers naissent aujourd'hui au jour de ta naissance,
Pour une fois le hasard tombe avec élégance.

Arijan, ne pars pas trop vite stp. À force, il ne reste plus beaucoup de Parisiens, avec qui est-ce que je vivrai ces aventures nocturnes dans le Ve maintenant ? Je vais me retrouver seul à tenir la boutique ... même si je ne pourrai jamais en vouloir à Emma.

[2]Comme il serait néanmoins un peu insultant, compte tenu de tes antécédents, que je te prive d'un jeu de piste pour les retrouver, je te donne seulement ceci : les réponses sont toutes quelque part dans cette thèse, mais, contrairement à ces remerciements, elles sont en anglais. Ce que tu cherches n'est pas leur emplacement, mais combien de fois tu les retrouves. Tu devrais avoir besoin de six souvenirs : ce que tu fais avec à peu près tout ; ce que tu égares avec une constance admirable quand tu veux rentrer chez toi ; ce que Claire a fini par accepter après avoir couru dans Montmartre ; ce qu'elle est devenue pour toi le 11 juillet dernier ; ce que nous avons joué pour vous ce jour-là avec Thibaud ; et ce que nous serons bientôt tous les deux pour Alex. Remets les six nombres dans cet ordre. Si tu sais où est mon site ; ajoute-les simplement après "/perdu/".

Pour les autres, ne m'en tenez pas rigueur, je ne peux pas continuer à faire des bêtises indéfiniment. Je remercie plus sobrement Perdig dont l'intelligence a toujours été un moteur d'innovation pour moi, Gus pour le courage de supporter Perdig et pour avoir dégoté la meilleure Luberonnaise possible, Zak pour prendre tant soin de Thibaud et pour toujours ouvrir tes portes à un pauvre étudiant de passage à Lyon, Clément pour ton peps constant et ta pédagogie dans le fonctionnement d'applications de rencontre, Olive pour ta capacité à basculer d'un sujet très sérieux à une bêtise en moins de 2 secondes, Thomas pour avoir gardé la chemise depuis l'école ... mais pas les cheveux, Spoutnik Wero pour m'avoir, un jour, pété le nez, et à Léon pour tous ces verres qu'on n'aura jamais pris alors qu'on bossait à 100m l'un de l'autre, à Marguerite pour les envois de banger baroques et de discussions toujours aussi quali. Enfin, comment ne pas remercier Corinne pour son accueil lors de mes trop rares passages en contrées albigeoises, son sourire et sa bonne humeur communicative. Ton dévouement constant envers les étudiants a beaucoup façonné la manière dont j'ai appris à interagir avec eux une fois passé de l'autre côté et je t'en remercie et suis toujours heureux de savoir que notre tableau trône toujours dans ton bureau dix ans plus tard.

Si ces remerciements sont aussi extensifs, c'est en partie à cause des membres de la famille Saka. L'initiation aux "beaux mails" de Maël et Dollon y est, j'en suis sûr, pour beaucoup. Un grand merci aux parrains, co-fillots et fillots (j'ai perdu le suivi des générations mais RDV à l'ABC), dont certains m'ont montré l'exemple de la thèse, et d'autres simplement soutenu - à leur manière - pendant mes corrections de copie.

Une expédition d'Albigeois s'est un jour envolée pour de nouvelles aventures québécoises, et je souhaite leur adresser un chaleureux merci à eux ainsi qu'aux anciens colocs. Clémence, pour avoir installé les chips au vinaigre de manière pérenne, sans que plus personne ne tique, Loulou pour ton soutien constant au karaoké et Adri pour ces quiz qui méritent à eux seuls un doctorat, Nina et Greg pour les dîners de darons et vos aventures que j'ai suivi beaucoup trop à distance, Caps pour tes montages plus que douteux et Élise pour avoir fait de Caps un troubadour, Paul pour ces dîners presque parfaits et ta recette de beurre qui m'aura convaincu que tu étais finalement vraiment cuisto, Zachy pour tes histoires amoureuses quotidiennes et ton kébab à la télé, Nono parce que premier gaou n'est pas gaou.

L'Enlise s'est un peu éparpillée mais finit ponctuellement par se retrouver. Albert, Hélène t'a donné une bonne raison et c'est toujours un plaisir de passer du temps avec vous trois, Adri pour les souvenirs de nouvel an et l'exportation de la recette de Caïpi, et bien sûr, je pense à toi Simon, pour les surprises à Mogador. Manon, tu avais ta place à bien des endroits dans ces remerciements mais je pense que l'Enlise ne te va pas trop mal. Merci à toi d'avoir été successivement ma voisine, puis ma camarade d'infortune de mémoire d'éco que nous découvrions, pour finir par être mon aînée de thèse. Tes conseils et ton exemple m'ont fait un bien fou, et je regrette de n'avoir pas eu l'occasion de descendre te voir à Nice. Reviens-nous vite !

C'est à Lyon que Benoît aura définitivement changé ma vie sans le savoir, en me suggérant, pendant un cours auquel nous ne prêtions pas particulièrement attention, de candidater au

poste d'assistant de recherche au LIEPP, qui aura déclenché la spirale infernale de la thèse. Tu as étrangement mis un sacré bazar dans ma vie, et je t'en remercie (des bisous à Alphonse !). Axelle pour ces (trop peu fréquents) verres parisiens et ces réalisations improbables d'amies en commun, Lucas pour toujours être partant pour festoyer, et Grégoire pour me rappeler que changer complètement de vie n'est pas impossible.

À Assas, la promotion Sauzon m'aura donné l'occasion de préparer mon entrée en thèse avec Chloé évidemment, la plus smart de tous (il aura fallu un petit temps avant qu'on ne réalise qu'on habitait dans la même rue), Gabriel toujours partant pour boire un verre, et Luc qui partage, à peu de choses près, la même vie que moi et pourrait être mon contrefactuel du groupe "sans thèse". J'ai évidemment une pensée pour toi Illan, ton écoute, tes conseils, et nos idées fantasques de papiers m'ont beaucoup aidé à me rassurer sur la direction que je prenais. J'espère qu'on réalisera un jour ces projets de papiers, pourquoi pas avec un statut un peu plus stable ? Je touche du bois.

Il existe, rue du Château des Rentiers, deux personnages peu recommandables qui m'accompagnent pourtant depuis vingt ans. Elsie, je n'aurais jamais dit que tu finirais tes études avant moi, et certainement pas pour une spécialité qui ne me sera d'aucune utilité avant encore quelques décennies. Philippe, je suis curieux de voir comment Nina calmera ton hyperactivité et j'espère que l'on continuera à avoir des discussions aussi douteuses que passionnées à Londres ou à Paris.

C'est souvent la tête bien pleine que j'arrivais à la Sirène les mardis soir pour l'orchestre, et la tête toujours aussi pleine, de musique cette fois-ci, que j'en ressortais. Je tiens à remercier le pupitre de V2 de l'orchestre Coalescence qui m'a définitivement convaincu que les V2 étaient vraiment le meilleur pupitre d'orchestre. Les Titines ont joué un rôle particulier. Je tiens à remercier Éléonore pour nos discussions féminisme et rupture de contrat de travail, Aliénor pour me rappeler que l'amour n'a pas d'âge, et Danica pour me rappeler que la lutte ne se gagne jamais assis sur une chaise. Pumpaaaaaj ! Enfin, Bertrand, je te dois beaucoup. Depuis cette rencontre inopinée au pupitre du Philharmonique de Jeunesse de Montréal, puis de l'orchestre de l'Isle, tu as été un compagnon de pupitre idéal et nous avons partagé beaucoup de choses, heureuses comme moins. J'espère que la fin de cette thèse s'accompagnera rapidement de ton retour au pupitre ! Les V2 ont besoin de toi. Paul, je suis heureux d'avoir fait ta rencontre, merci pour les échanges intimes que l'on a pu avoir ces dernières années et sache que je compte bien te renvoyer l'ascenseur pour tous ces concerts !
Je tiens également à remercier Sylvie, dont notre vouvoiement constant malgré toutes ces années est une marque, contre-intuitive, de notre affection croissante. J'espère que nous aurons plus de temps pour ces petits verres tardifs et restaurants kurdes.

Georges, tu as beau venir de Boulogne, je te pardonne. J'espère que les années à venir entraîneront des dîners plus fréquents à parler d'économie, de chemises ... douteuses, et de vin. Et tout ça, c'est aussi en partie grâce à Inno, alors merci à elle !

Elisa, tu es arrivée dans nos vies comme un épisode de “Love is blind, habibi”: n'importe comment. Nos escapades à l'Engrenage (pas fan de la nouvelle DA d'ailleurs, faut faire remonter à Alex) sont toujours une nouvelle aventure. J'attends avec impatience le jour où tu déménageras enfin de ce four qui ruine des bouteilles de Château Margaux mais, en attendant, ta chambre est toujours prête (et arrête de te battre avec Antoine stp). Hâte que tu nous ramènes un tigre.
Adeline, notre dette de cadeaux envers vous ne cesse de s'agrandir, et j'espère que la venue d'Auguste nous permettra de combler rapidement ce déficit. Orléans n'est pas si loin, et je compte bien apprendre de nouvelles recettes de cocktail dans ce bar extérieur après d'Arnaud, jeune docteur en mixologie.
Taboni, tu es l'Alexandre préféré de Charlotte, et, je n'arrive pas à croire que je vais écrire cela, je te remercie pour ton humour ... décalé et pour mettre un peu de fumée dans nos vies. Je remercie surtout Sandra pour la patience dont elle fait preuve à ton égard. Elle est la preuve que le cœur a vraiment ses raisons.
Marie pour m'avoir montré que les belges étaient définitivement bien plus créatifs que les français en améliorant les coupe du monde en douches picole.
Giulia, even though we haven't had much time to see each other over the last few years, I couldn't be happier for you and Becky, and I can't wait to celebrate with you and pop round to see you both ... but more importantly, to see Blue and remind myself why I'll always prefer cats.
À toutes et tous, merci de prendre autant soin de Charlotte.

De Boulogne à Pasteur, de Mer à Bécon et de Corse à Vaalwater, j'ai toujours reçu un accueil chaleureux chez Isabelle et Pierre. Si, au fil du temps, mes réserves de Coureur des Bois se sont quelque peu amenuisées, celles de rhum et d'Amarula n'ont, elles, cessé de s'étoffer. Je ne me suis pas encore complètement transformé en “Birder” mais je suis heureux d'avoir pu cocher cette étape avec vous. Un grand merci à vous, ainsi qu'à Serge, Thomas, Gwendo et les deux dinos pour avoir égayé mes déjeuners dominicaux. J'ai évidemment une pensée émue pour Micheline.

J'ai une dette immense envers ma famille. Votre soutien a été constant, discret quand il le fallait, et toujours là quand j'en avais besoin. À mes parents d'abord, qui ont toujours fait en sorte que je n'aie pas à penser à autre chose qu'à mes études, me permettre de me reposer à Thury comme à Safi, je mesure aujourd'hui pleinement la chance que j'ai eue grâce à vous. Guillaume, tu es devenu au fil des ans, le pilier de cette famille et je te dois beaucoup. Mathieu, je sais qu'on ne le dit pas suffisamment mais je suis fier de tes progrès et je suis heureux de pouvoir passer plus de temps avec toi. Papi, tu as toujours fait en sorte de me titiller là où ça dérangeait (à coups de bonnes bouteilles de vin quand même), de me mettre un bouquin entre les mains. Je vous dois énormément, à mamie et toi. J'ai une pensée pour Marylou, et nos soirées canadiennes, souvent pleines de confidences. Enfin, j'aimerais remercier les grandes-tantes et cousins, cousines et les Bretons de Concarneau/Trégunc et en profiter pour m'excuser de n'avoir pas profité autant de la longère que j'aurais aimé, ce n'est j'espère que partie remise, en Bretagne ou ailleurs. Merci également à Didier de nous avoir fait l'honneur d'un rarissime passage parisien, il semblerait qu'un jardin puisse accomplir des miracles !
J'ai également une pensée pour ma famille marocaine. La dernière fois que je vous ai vus, je

rentrais à peine en thèse. Vos photos et pensées pour moi tout au long de ces années m'ont toujours fait chaud au cœur. Tabti bien sûr, dont la voix m'apporte toujours du réconfort, Fatouche, pour ta combativité, tes conseils cinéma et tes photos de chats, Said et Latifa pour votre bonne humeur et Hamidou pour ton soutien. Félicitations à Zaynab et Zizou pour votre réussite, et pour toi Ziyad, on espère te voir bientôt à Paris pour tes études.
Je dédie cette thèse à Martine, ma grand-mère, et Dadda Amghar, mon grand-père, qui ont toujours insisté pour mettre l'éducation au-dessus de tout.

Ce dernier paragraphe de remerciements va à Maurice et Charlotte[3]. Je commence donc par toi Maurice[4], et te remercie d'avoir égayé notre quotidien (et maintenant celui de nos voisins) à coups de câlins et de bruits de clochette.
Enfin, Charlotte, je suis heureux de rendre cette thèse au seuil de ma quatorzième année passée à tes côtés, et, fait inédit, la première où je n'aurai techniquement plus de carte étudiante. Tu es, depuis toutes ces années, mon point d'ancrage, tu m'as soutenu à chaque étape, dans chacun de mes choix, même lorsqu'ils m'emmenaient à l'autre bout de France, puis de la planète, puis simplement un peu plus bas dans la distribution des salaires. Tu as toujours su trouver la patience, les mots pour apaiser mes craintes, et me rappeler, quand il le fallait, de prendre un peu de temps pour moi. Après tant d'années, et presque autant de détours, je mesure toute la chance que j'ai de t'avoir à mes côtés. Je ne sais pas encore ce que cette nouvelle vie sans carte étudiante nous réserve, mais j'ai hâte de continuer à la découvrir à tes côtés.

À Ivry-sur-Seine,
Le 16 août 2026.

[3] Pour ne pas créer de tension, l'ordre d'apparition dans ce paragraphe a été attribué aléatoirement à l'aide de l'*Author Randomization Tool* de l'*American Economic Association* et peut être retrouvé ici.

[4] J'ai cherché un précédent à de tels remerciements pour un chat et suis heureux d'en avoir trouvé un auprès de Mathilde Lesueur, qu'elle en soit encore remerciée.

Résumé

Les programmes publics peuvent être formellement accessibles sans pour autant atteindre effectivement les populations auxquelles ils sont destinés. Cette thèse étudie en France trois dimensions de cette relation : l'engagement, l'allocation et l'intensité de l'intervention, et l'accès aux services.

Le premier chapitre étudie un programme proposant des ateliers de langue et d'intégration aux parents immigrés au sein des établissements scolaires de leurs enfants. Le programme n'améliore pas leur niveau de français, mais accroît leur participation aux activités scolaires et leur estime de soi. Leurs enfants progressent également de 0,25 à 0,30 écart-type aux évaluations de français, suggérant que l'engagement avec l'école peut constituer un canal efficace d'intégration.

Le deuxième chapitre porte sur les centres éducatifs fermés, créés comme une alternative à l'incarcération des mineurs. Parmi les mineurs pour lesquels un CEF a été recommandé, le placement se substitue principalement à des mesures de suivi en milieu ouvert plutôt qu'à la détention et augmente de 8 à 10 points de pourcentage la probabilité d'une détention dans les douze mois. Les résultats suggèrent que la révocation de la mesure judiciaire portant le placement constitue un mécanisme important.

Le troisième chapitre étudie l'accès des réfugiés aux soins de santé mentale. Dans une expérimentation randomisée menée auprès de 1 135 réfugiés, seule la combinaison d'un échange avec un pair-aidant et la mise à disposition d'une ligne téléphonique multilingue facilitant la prise de rendez-vous augmente le recours aux soins, de 6,2% à 13,7%, malgré un recours qui demeure faible.

Les résultats de cette thèse soulignent que la capacité des institutions à atteindre leurs publics dépend non seulement de ce que les programmes proposent, mais aussi de la manière dont sont organisés le contact, l'accompagnement et l'intervention.

Mots clés : Économie publique, Migration, Justice des mineurs, Santé

Abstract

Public programmes can be formally accessible without effectively reaching the populations they are intended to serve. This thesis studies three dimensions of this relationship in France: engagement, allocation and intensity of intervention, and access to services.

The first chapter studies a programme offering language and integration workshops to immigrant parents inside their children's schools. The programme does not improve parents' French, but increases their participation in school activities and self-esteem. Their children also gain 0.25 to 0.30 standard deviations in French test scores, suggesting engagement with the school as an efficient channel of integration.

The second chapter studies closed educational centres, created as an alternative to juvenile incarceration. Among minors recommended for a CEF, placement mostly displaces community supervision rather than detention and raises the probability of detention within twelve months by 8 to 10 percentage points. The results suggest that revocation of the judicial order carrying the placement is an important channel.

The third chapter studies refugees' access to mental health care. In a randomised experiment among 1,135 refugees, only the combination of a peer-helper discussion and a multilingual helpline facilitating appointment booking increases engagement with care, from 6.2% to 13.7%, despite take-up remaining low.

Overall, the findings highlight that institutional reach depends not only on what programmes offer, but also on how contact, assistance and intervention are organised.

Keywords: Public Economics, Migration, Juvenile Justice, Health

Contents

Remerciements vii

Résumé xix

Abstract xxi

Contents xxiii

List of Figures xxvii

List of Tables xxix

General Introduction 1

1 Opening Schools to Parents to Promote Migrants' Inclusion and their Children's Achievement 17

1 Introduction . . . 19
2 Context and Programme Description . . . 22
2.1 Newly Arrived Immigrants' Integration Pathway in France . . . 22
2.2 The OEPRE Programme . . . 24
3 Research Design . . . 26
3.1 Pilot and Qualitative Study . . . 26
3.2 Sampling Strategy . . . 28
3.3 Timeline . . . 31
4 Data and Descriptive Statistics . . . 31
4.1 Data Sources . . . 31
4.2 Description of Parents and Schools . . . 32
4.3 Outcomes of Interest . . . 34
5 Estimation Strategy . . . 35
5.1 Balance Checks between Schools and Parents . . . 35
5.2 Intention-to-Treat Estimates . . . 36
5.3 Effects on Programme Participants . . . 37
6 Results . . . 39
6.1 Intention-to-Treat Estimates . . . 39
6.2 Effects on Programme Participants . . . 42
7 Conclusion . . . 47
References . . . 49
Appendix . . . 51
A Additional Tables and Figures . . . 51
B Construction of the Variables and Indices . . . 57
B.1 Outcomes . . . 57

B.2 Control variables . . . 62
C Main Estimates . . . 64

2 Educate and Punish: Allocation, Displacement, and the Consequences of an Intermediate Juvenile Sanction 71
1 Introduction . . . 73
2 Institutional Setting . . . 78
2.1 The Closed Educational Centre . . . 78
2.2 The Other Responses Available to the Court . . . 80
2.3 Allocation of Minors to a Closed Educational Centre . . . 81
3 Data . . . 82
3.1 Data Sources and Sample Selection . . . 82
3.2 Variables of Interest . . . 85
3.3 Descriptive Statistics . . . 86
4 Empirical Strategy . . . 88
4.1 Identification Challenge . . . 88
4.2 Defining the counterfactual . . . 89
4.3 Documenting the allocation to a closed educational centre . . . 91
4.4 Estimating the effects of placement in a closed educational centre . . . 92
5 Results . . . 95
5.1 The decision process of the placement . . . 95
5.2 The effects of the placement . . . 99
6 Mechanisms . . . 110
7 Conclusion . . . 116
References . . . 118
Appendix . . . 121
A The French Juvenile Justice System . . . 121
A.1 Judicial Process and the Severity Ladder of Sanctions . . . 121
A.2 Judicial Measures Available to a Juvenile Court . . . 124
A.3 Geographic Roll-out of the Network . . . 127
B Data Coverage and Measurement . . . 130
B.1 Prescriber's Department Coverage (2004-2006) . . . 130
B.2 Detention: Data Availability (2010-2012) . . . 133
B.3 Variable Definitions . . . 134
B.4 Baseline Characteristics of the Analytical Cohorts . . . 136
B.5 Selection into the Report-Linked Sample . . . 136
C Comparison Group, Balance and Sensitivity . . . 139
D The Placement Decision . . . 150
E Additional Estimates . . . 156
F Estimates Anchored at the End of the Measure . . . 159
G Additional Mechanism Evidence . . . 163

3 The Impact of Lifting Barriers to Mental Health Care for Refugees 169
1 Introduction . . . 171
2 Institutional Context and Interventions . . . 175
2.1 Refugees in France and the AGIR programme . . . 175
2.2 The Interventions . . . 177
2.3 Theory of Change . . . 179
3 Design of the Experiment and Data . . . 180
3.1 Sample Construction . . . 181

3.2 Experimental Protocol . . . 183
3.3 Data sources . . . 185
3.4 Variables of interest . . . 187
4 Sample description . . . 189
4.1 Refugees' Health Profile . . . 189
4.2 Refugees' Economic and Housing Integration . . . 192
5 Empirical Strategy . . . 193
5.1 Balance between Groups . . . 193
5.2 Intention-to-Treat estimates . . . 193
5.3 2SLS estimates . . . 195
5.4 Heterogeneous Effects . . . 195
6 Results . . . 196
6.1 Results on take-up, engagement with mental healthcare and hospital visits 196
6.2 Results on housing, employment and exit of the programme . . . 204
7 Heterogeneous effects . . . 208
8 Conclusion . . . 210
References . . . 212
Appendix . . . 217
A Implementation of the Intervention . . . 217
A.1 Intervention materials . . . 217
A.2 Sample construction and selection into the study . . . 221
B Data and Variable Construction . . . 224
B.1 Outcomes . . . 224
B.2 Control Variables . . . 229
B.3 Baseline Mental-Health Instruments . . . 231
B.3.1 International Trauma Questionnaire (ITQ): PTSD symptoms . . . 231
B.3.2 ITQ: PTSD functional impairment . . . 232
B.3.3 ITQ: Disturbances in Self-Organisation (DSO) symptoms . . . 232
B.3.4 ITQ: DSO functional impairment (not collected at baseline) . . . 233
B.3.5 ITQ: Scoring rules . . . 233
B.3.6 Depression screen (IDQ items 1–2, equivalent to PHQ-2) . . . 233
B.4 Baseline Mental Health . . . 234
C Balance and Baseline Characteristics . . . 238
D Main Estimates by Outcome Family . . . 243
D.1 Healthcare utilisation . . . 243
D.2 Exit from the AGIR programme . . . 247
D.3 Employment and housing trajectories . . . 248
D.4 Robustness . . . 251
D.5 Effects on the Treated . . . 256
E Heterogeneous Effects . . . 261
E.1 Effects by Gender . . . 261
E.2 Counterfactual Peer Assignment . . . 269

List of Figures

General Introduction **1**

1 Opening Schools to Parents to Promote Migrants' Inclusion and their Children's Achievement **17**
1.1 Sample Construction . . . 28
1.2 Geographic Distribution of Study Schools, 2022–2023 . . . 30
1.3 Study Timeline . . . 31
1.4 ITT Effects on parental language proficiency and engagement in schooling . . . 40
1.5 ITT Effects on parental political, and economic outcomes . . . 42
1.6 ITT effects on children's test scores . . . 43
1.7 Effects of OEPRE Participation on parental language proficiency and engagement in schooling . . . 44
1.8 Effects of OEPRE participation on parental political, social, and economic outcomes 45
1.9 Effects of OEPRE participation on children's test scores . . . 46
A1 Geographic Distribution of OEPRE Schools, 2019–2024 . . . 51
A2 Balancing between treated and non-treated parents for parental outcomes, before and after CBPS and entropy balancing score-based reweighting . . . 52
A3 Balancing between treated and non-treated parents for children's outcomes, before and after CBPS and entropy balancing score-based reweighting . . . 53

2 Educate and Punish: Allocation, Displacement, and the Consequences of an Intermediate Juvenile Sanction **71**
2.1 Cumulative roll-out of CEFs in France, 2003-2022. . . . 79
2.2 Within-stay events and exit pattern of the closed educational centre stay . . . 113
2.3 Cumulative incidence of post-spell events by effective stay length . . . 115
A1 Simplified representation of the French juvenile penal pathway . . . 122
A2 Severity ladder of placement decisions and the CEF as accessory to a coercive vehicle. . . . 123
A3 Network capacity and national occupancy of CEFs, 2004-2024. . . . 127
A4 Geographic distribution of CEFs across French departments in 2022. . . . 128
A5 Year-by-year geographic expansion of CEFs across French departments, 2003-2022. 129
B1 Prescriber department coverage by imputation level, 2004-2012 . . . 132
B2 Detention data coverage: raw spell counts by year . . . 133
C1 Distribution of the placement propensity score in the two arms. . . . 139
C2 Standardised covariate differences before and after reweighting. . . . 141
D1 Persistence of facility intake profiles across years . . . 154
G1 Survival of the judicial supervision order with and without an attached placement 164

3 The Impact of Lifting Barriers to Mental Health Care for Refugees 169
3.1 Theory of change . . . 179
3.2 Flow diagram of participants (FTDA sample) . . . 181
3.3 Breakdown of nationalities . . . 190
3.4 Baseline PTSD prevalence, by subgroup . . . 191
3.5 Baseline prevalence of unemployment, by subgroup . . . 192
3.6 Standardised differences across treatment arms . . . 194
3.7 Effect on any post-treatment signal, by treatment arm . . . 197
3.8 Effects on helpline outcomes, by treatment arm . . . 199
3.9 Effects on partner-structure outcomes, by treatment arm . . . 201
3.10 Effects on hospital outcomes, by treatment arm . . . 202
3.11 Probability of still being enrolled in AGIR, by treatment arm . . . 205
3.12 Share of refugees active and in stable housing, by week relative to treatment . . . 207
A1 Rollout and inclusions in the AGIR programme . . . 217
A2 RISE Project Logo . . . 217
A3 Information Leaflet . . . 218
A4 Helpline Label . . . 219
A5 Video presentation of the project . . . 219
A6 Covariate Balance Between Individuals Seen and Not Seen by RISE . . . 221
A7 Consent Rate Over Time . . . 222
A8 Cumulative Inclusion of Refugees in RISE . . . 222
A9 Covariate Balance by Consent Status Among Individuals Seen by RISE . . . 223
B1 Distribution of ITQ PTSD Symptom Scores at Baseline . . . 234
B2 Baseline depression-screen positive rate, by subgroup . . . 237
D1 Effects on employment and housing trajectories, by treatment arm . . . 250

List of Tables

General Introduction **1**

1 Opening Schools to Parents to Promote Migrants' Inclusion and their Children's Achievement **17**
1.1 First Stage: Effect of School Assignment on Workshop Participation 37
1.2 ITT estimates on parental participation in school events 41
A1 Baseline Characteristics of Sampled Parents 54
A2 Balance Check: Treatment vs Control School Characteristics 55
A3 Balance Checks: Parent Characteristics 56
C1 ITT estimates on parental language proficiency and engagement in schooling . . 64
C2 ITT estimates on parental civic, social, and economic outcomes 65
C3 ITT estimates on children's test scores, with school covariates 66
C4 ATT estimates on parental language proficiency and engagement in schooling . . 67
C5 ATT estimates on parental civic, social, and economic outcomes 68
C6 Effects of OEPRE participation on children's test scores, with school covariates . 69

2 Educate and Punish: Allocation, Displacement, and the Consequences of an Intermediate Juvenile Sanction **71**
2.1 Comparison of residential modalities for juvenile offenders in France 81
2.2 Offence shares by response family, each minor weighted equally 87
2.3 Judicial responses by prescribing authority 95
2.4 What predicts the judicial response: recorded trajectory, nature of the offence, and the educator's recommendation 96
2.5 Similarity of co-entrants against a timing-permutation placebo 99
2.6 Destination of minors recommended for a CEF but not placed in one 100
2.7 Measures recommended in the report attached to minors placed in a closed educational centre 101
2.8 Effect of a CEF placement within the recommended population, outcomes anchored at the decision (PARCOURS, 2021-2024) 102
2.9 Placement effect on the full report-linked sample, with and without recommendation controls (PARCOURS, 2021-2024) 103
2.10 Heterogeneity of the placement effect by recommendation status (PARCOURS, 2021-2024) 104
2.11 Placement effect on the subsample with a recorded offence nature, outcomes anchored at the decision (PARCOURS, 2021-2024) 105
2.12 Effect of a first CEF placement by counterfactual type, twelve and eighteen months from the decision 106

2.13 Delay between the end of the indexed spell and the first subsequent detention episode . . . 111
2.14 In-stay incidents: closed educational centre against the reweighted residential counterfactual . . . 112
2.15 Decomposition of post-spell detention by within-stay incident profile . . . 114
A1 Main judicial measures in the French juvenile justice system . . . 124
B1 Coverage of prescriber department by imputation level . . . 131
B2 Prescriber department vs structure department match rates by measure family . 131
B3 Detention data coverage by year . . . 133
B4 Duration of the indexed spell and age at its end, by arm and control type . . . 134
B5 Baseline characteristics of minors at first placement, first closed educational centre, and first detention . . . 136
B6 Measures with and without a linked social-educational report (PARCOURS, 2021-2024) . . . 137
B7 Subsequent judicial events of closed centre placements, with and without a linked report . . . 138
C1 Covariate balance before and after entropy-balancing reweighting, common-support sample . . . 140
C2 Composition of the comparison group on common support . . . 142
C3 Selection-on-unobservables statistics for the main comparison . . . 142
C4 Strength an unobserved factor would need to overturn the estimate, report-linked sample and recommended population . . . 143
C5 Placebo balance on pre-determined variables and on outcomes measured before the decision . . . 144
C6 Sensitivity of the detention estimates to the imbalanced pre-determined variables 144
C7 Where the adjustment comes from: raw, covariate-adjusted, inverse-probability weighted and entropy-balanced estimates . . . 145
C8 Common-support rules and resulting analytical samples . . . 145
C9 Decision-anchored effect within the recommended population, alternative fixed effects . . . 146
C10 Effect of a first CEF placement, control decisions down-weighted by minor, twelve and eighteen months from the decision . . . 147
C11 Effect of a first CEF placement with pre-decision outcomes in the conditioning set 148
C12 Concentration of the entropy-balancing weights, by comparison panel . . . 149
D1 Judicial response as a choice between alternatives and as a position on a single severity dimension . . . 150
D2 Rank of the receiving facility in the distance ordering of centres . . . 151
D3 Matching of minors to centres before and after the 2016 shift to a national admission procedure . . . 151
D4 Occupancy of the skipped nearer centres at the month of the placement . . . 152
D5 Local availability of a place and the probability that a recommended minor is actually placed . . . 152
D6 Facilities ranked by the severity of the profiles they admit . . . 153
D7 Year-to-year transitions of facilities across thirds of intake severity . . . 153
D8 Rank and distance of the receiving facility by severity of the minor's history . . . 155
E1 Effect of a first CEF placement by counterfactual type, comparison-specific support 156
E2 Effect of a first CEF placement within the legal vehicle of the placement . . . 157
E3 Effect of a first CEF placement by counterfactual type, retaining single-record placements . . . 158

F1 Effect of a CEF placement within the recommended population (PARCOURS, 2021-2024) . . . 159
F2 Effect of a first CEF placement by counterfactual type, common-support sample 160
F3 Effect of a CEF placement by length of the stay, within the recommended population (PARCOURS, 2021-2024) . . . 161
F4 Placement effect on the subsample with a recorded offence nature (PARCOURS, 2021-2024) . . . 162
G1 Custody that closes the measure it interrupts, by type of measure . . . 163
G2 Effect of a CEF placement by length of the stay, decision-anchored outcomes . . . 165
G3 Sector and programme-quality differentials in the effect of a first CEF placement 166
G4 Effect of a first CEF placement by location of the facility relative to the prescriber 167

3 The Impact of Lifting Barriers to Mental Health Care for Refugees 169
3.1 Peer-assignment effect by (counterfactual) mediator and arm . . . 210
B1 Distribution of AGIR exit motives across treatment arms . . . 227
B2 Exit categories validated against the collected trajectories . . . 228
B3 Exit composition by arm (descriptive) . . . 228
B4 ITQ Item Distribution and Diagnostic Construction . . . 235
B5 Evaluability per ITQ Diagnostic Component . . . 236
B6 PTSD Prevalence by Missingness Group (Diagnosable Cases Only) . . . 236
C1 Balance tests on baseline characteristics across randomised groups . . . 238
C2 Descriptive Statistics by Treatment Group: Socioeconomic Characteristics at Baseline . . . 240
C3 Descriptive Statistics by Treatment Group: Health and Mental-Health Status at Baseline . . . 242
D1 ITT Effects on any-Signal Outcomes . . . 243
D2 ITT Effects on Hospital Outcomes . . . 244
D3 ITT Effects on Partner-Structure Outcomes . . . 245
D4 ITT Effects on Helpline Outcomes . . . 246
D5 ITT Effects on Exit Outcomes . . . 247
D6 ITT Effects on Trajectory (Employment and Housing) Outcomes . . . 248
D7 ITT Effects on Trajectory Intensive Margins Outcomes . . . 249
D8 ITT Effects on any-Signal (with Social-Worker FE) Outcomes . . . 251
D9 ITT Effects on Hospital (with Social-Worker FE) Outcomes . . . 252
D10 ITT Effects on Partner-Structure (with Social-Worker FE) Outcomes . . . 253
D11 ITT Effects on Helpline (with Social-Worker FE) Outcomes . . . 254
D12 ITT Effects on Exit (with Social-Worker FE) Outcomes . . . 255
D13 2SLS estimates: effect of any post-treatment healthcare contact on trajectory outcomes . . . 256
D14 2SLS estimates: effect of any post-treatment healthcare contact on intensive-margin trajectory outcomes . . . 257
D15 2SLS estimates: effect of calling the helpline on trajectory outcomes (LATE) . . . 258
D16 2SLS estimates: effect of calling the helpline on intensive-margin trajectory outcomes (LATE) . . . 259
D17 2SLS estimates: effect of any post-treatment healthcare contact on AGIR exit outcomes . . . 260
E1 Heterogeneous Effects by Gender: Any Post-Treatment Signal . . . 261
E2 Heterogeneous Effects by Gender: Hospital Outcomes . . . 262
E3 Heterogeneous Effects by Gender: Hospital Outcomes (with covariates) . . . 263

E4 Heterogeneous Effects by Gender: Partner-Structure Outcomes 264
E5 Heterogeneous Effects by Gender: Helpline Outcomes 265
E6 Heterogeneous Effects by Gender: AGIR Exit Outcomes 266
E7 Heterogeneous Effects by Gender: Trajectory Outcomes 267
E8 Heterogeneous Effects by Gender: Intensive-Margin Trajectory Outcomes 268
E9 Counterfactual peer assignment: validation on treated refugees 269
E10 Actual (peer arms) and counterfactual (other arms) mediator assignment 270
E11 Peer-assignment effect by (counterfactual) mediator 270

General Introduction

Public policy does not operate through formal eligibility or service provision alone. Its effects depend on whether institutions establish an effective relationship with the populations they are meant to serve, on the form this relationship takes, and on whether the services offered are actually used. A large literature on the take-up of social benefits documents a persistent gap between eligibility and take-up: across countries and programmes, a substantial share of those entitled to a benefit do not claim it (Currie, 2004; Ko and Moffitt, 2022). The gap can be large, even for generous programmes. In the United States, around a quarter of those eligible for the largest means-tested cash transfer do not claim it, forgoing on average the equivalent of a month of income (Bhargava and Manoli, 2015). Understanding this gap between formal eligibility and effective access requires looking at the frictions that stand between individuals and public programmes.

Informational, administrative and psychological frictions can stand between a programme and the population it is designed to serve: people do not know that they are eligible, forms are long, procedures are confusing, offices are far, and claiming can feel demeaning (Herd and Moynihan, 2025). However, these frictions are not fully exogenous, and are partly shaped by policy design. Complexity can arise from effort to screen out ineligible applicants, so decisions intended to improve targeting can themselves make participation more difficult (Kleven and Kopczuk, 2011). Programme design therefore shapes not only overall take-up, but also who is ultimately reached.

These barriers do not affect everyone equally either. When an American food-stamp experiment offered information and application assistance to eligible elderly non-participants, enrolment tripled, but the individuals induced to enrol were richer and healthier than existing recipients, and frictions appeared greatest among those with the highest needs(Finkelstein and Notowidigdo, 2019). When offices assisting with disability applications close, those who give up are disproportionately less educated and have moderately severe conditions, precisely the individuals the programme is intended to support (Deshpande and Li, 2019). Some barriers are relational rather than administrative. Black men randomly assigned to Black doctors take up substantially more preventive care after meeting them (Alsan et al., 2019), and the disclosure of the Tuskegee syphilis study, in which Black men were deliberately deceived about their diagnosis and denied treatment so that researchers could observe the progression of the disease, had lasting effects on older Black men's use of medical care and trust in the medical system (Alsan and Wanamaker, 2018). Institutional identity and past treatment can therefore shape whether formal access translates into actual use.

The challenge is not always a question of insufficient contact. For populations already closely involved with public institutions, policy design also determines the form and intensity of that relationship, and which interventions are chosen among the available alternatives. In this thesis, hard-to-reach therefore does not describe an intrinsic characteristic of individuals, but rather a relationship between individuals and institutions in which policy design shapes whether contact occurs, how people engage with institutions once contact is established, and what form intervention takes.

This thesis studies three margins of this relationship. The first is *engagement*: when an institution is already present in people's lives, can it deepen their engagement with it, with effects beyond its immediate role? The second is *allocation and intensity*: when institutional contact is already intensive, which intervention does the state choose, what does that intervention actually replace, and with what consequences? The third is *access*: when a service is formally available, which barriers prevent those who need it from using it, and how can they be lifted? The first chapter asks whether engagement with schools can foster immigrant parents' integration. The second studies young offenders already deeply involved with the justice system and asks how they are allocated across interventions of different intensity, and with what consequences. The third turns to refugees with substantial mental health needs and asks what prevents formal access to care from translating into actual use.

The three chapters are set in France and examine different points of contact between populations and public institutions. Non-take-up is widespread in the French welfare state: a third the households eligible for the guaranteed minimum income do not claim it in any given quarter (DREES, 2022), while a recent nationwide experiment shows that assistance can increase take-up where information alone does not, and that those who stand to gain most from a meeting assessing the individual's rights are among the least likely to attend (Castell et al., 2025). This thesis studies three distinct institutional settings: workshops for immigrant parents on French language, the school system and Republican values, held inside their children's schools; an individualised support programme for refugees; and closed educational centres for juvenile offenders. Each chapter draws on administrative data or data collected in the field.

The first chapter, joint with Flore Gubert, Élise Huillery and Jean-Noël Senne, evaluates OEPRE, a programme that hosts language and integration workshops for immigrant parents inside their children's schools. Exploiting the fact that otherwise comparable schools do or do not host the programme in a given year, we find that the programme does not improve parents' French, but strengthens their engagement with the school: participation in school activities rises by 25% of a standard deviation, self-esteem by 20%, and the children of participating parents gain 25 to 30% of a standard deviation in French test scores, with no effect in mathematics. Integration policy usually treats language as the gateway; the chapter shows that stronger attachment to an institution can generate benefits even when language training does not.

The second chapter, solo-authored, studies closed educational centres (CEF), a residential measure of the French juvenile justice introduced in 2002 as an alternative to incarceration. Using the placement recommendation written by the duty educator before each hearing, I compare minors recommended for a CEF who were actually placed with recommended minors who were

not. For most of the minors it receives, the centre turns out to be an addition to community supervision rather than a substitute for detention. Placement raises the probability of a detention episode within twelve months by 8 to 10 percentage points among recommended minors, and the timing of subsequent detention is consistent with revocation of the order carrying the placement being an important channel. In this setting, what matters is therefore not whether the institution reaches the individual, but the form and intensity of the intervention once contact is already established.

The third chapter, joint with Flore Gubert, Marie-Caroline Saglio-Yatzimirsky, Hala Tobji and Jérôme Valette, presents the results of a randomised experiment among 1,135 refugees enrolled in France's main integration programme, and studies how to turn a free access to mental health care into actual contact with care. We randomly assigned refugees to four arms: a control group, a confidential discussion with a peer-helper that targets stigma and mistrust, a multilingual helpline that books appointments on the refugee's behalf and targets navigation in the healthcare system, and the combination of both interventions. Only their combination increases overall engagement with care, more than doubling the probability of any recorded contact, from 6.2% to 13.7%. We do not find a detectable increase in engagement from the peer-helper discussion or the helpline alone, suggesting complementarities between socio-cultural and practical barriers. Take-up nevertheless remains low even when both barriers are addressed jointly.

Taken together, these chapters show that the effects of public interventions depend not only on their content, but also on the relationship institutions build with the populations they serve. In the first chapter, bringing the programme into schools changes parents' engagement even though their French does not improve. In the third, neither the peer-helper discussion nor practical assistance *alone* is enough to increase overall engagement with care; the effect appears only when assistance is combined with an intervention addressing stigma and mistrust. Both chapters therefore point to dimensions of institutional engagement that go beyond the formal availability or content of a service. The juvenile justice chapter starts from a different point: young people are already deeply involved with the system. The relevant question is then which intervention they receive, what it replaces, and how it affects later trajectories. Across these settings, institutional reach emerges not as a binary question of whether a population is reached, but as a feature of how public policy organises contact, assistance and intervention.

The remainder of this introduction describes each chapter in more detail.

Chapter 1: Opening Schools to Parents to Promote Migrants' Inclusion and their Children's Achievement

Co-authored with Flore Gubert, Élise Huillery and Jean-Noël Senne

Over the last three decades, the share of immigrants in Europe and Northern America has nearly doubled, rising from around 8% in 1990 to roughly 14% in 2024 (Nations, 2024). How immigrant families integrate into their host society has therefore become a central policy question, and it

is typically studied through the lenses of language acquisition, citizenship, and labour market outcomes. Yet for many immigrant parents, and for mothers in particular, everyday contact with host-country institutions happens first and foremost through their children's school. Unlike stand-alone language or employment programmes, schools offer a recurring setting in which parents already have a reason to show up. Whether schools can serve as effective sites of integration for immigrant parents is the question this first chapter addresses.

Parental involvement in schooling is a key determinant of children's academic success (Hoover-Dempsey and Sandler, 1995; Avvisati et al., 2010; Kim, 2022), but immigrant parents are consistently less involved in their children's schooling than native-born parents (Guryan et al., 2008; Turney and Kao, 2009), and a persistent achievement gap separates students of immigrant background from their native peers, even net of socio-economic background and home language (OECD, 2025). Additionally, immigrant mothers are hard to reach through conventional integration programmes: lower language proficiency, childcare responsibilities and social norms limit their engagement with services located outside familiar spaces, even though comparative evidence suggests that targeting human capital policies towards women may maximise their long-run impact (Foged et al., 2024a). Therefore, can interventions engaging parents through the school setting rather than through language training alone help close some of this gap?

We study this question using the "Opening Schools to Parents for their Childrens' Achievement" programme (*Ouvrir l'École aux Parents pour la Réussite des Enfants* - OEPRE), a school-based integration programme funded by both the French Ministry of Education and of Home Affairs. OEPRE is part of the Republican integration pathway for newly arrived foreign nationals, and offers 60 to 120 hours of free training per school year, in groups of 8 to 15 parents, combining French courses that can lead to official language certification, an introduction to the French school system and the values of the French Republic. Sessions are held during or after school hours, mostly in primary and lower secondary schools, and are run by teachers of French as a foreign language or social workers. The programme now operates in more than 500 schools. Enrolment rose from about 14,000 parents in 2019-2020 to nearly 19,000 in 2022-2023, and roughly 90% of participants are women. Classes therefore bring together parents who share the same daily reality and are already anchored in the school through their children. This particularity allows us isolate the effect of a school-based, community-oriented intervention from that of language acquisition only.

Two preparatory studies shaped the evaluation. A pilot survey of 470 schools showed that hosting the workshop in the school attended by the parents' own children raises participation by 34 percentage points on a base of 21%. Qualitative work in 14 schools and with 37 participating parents highlighted the removal of an existing "linguistic shame", greater autonomy in interactions with teachers, and the development of social ties beyond family networks. These findings guided the survey design.

Because the ministry cannot open a workshop in every school with many allophone parents, otherwise similar schools host or do not host the programme in a given year, and our evaluation builds on this variation. Starting from 576 candidate schools, we matched each of the 156 schools hosting OEPRE in 2022-2023 to a comparable school without the programme that year using

pre-treatment school characteristics, local socio-economic conditions and previous achievement of children. We then interviewed 713 parents by phone in their preferred language, 448 in treatment schools and 265 in control schools, spread over 151 schools. Treatment is defined at the school by year level: parents whose children attended a school in a year in which it hosted OEPRE form the treatment group, whether or not they themselves participated.

The survey collects demographics, migration histories, education and residency status, together with the outcome modules, and we retrieve children's tests scores at the national evaluations in administrative data provided by the statistical department of the Ministry of Education. The population we reach is the one integration policy struggles with: 75% of respondents are women, with a low self-assessed French proficiency at arrival, 2.7 children on average, and almost half of respondents hold a precarious residency status. Schools are just as far from the national average, as 77% of treatment schools belong to priority education against 16% nationally.

Participation is voluntary and imperfect on both sides: 54% of surveyed parents participate when their children's school hosts the programme, against 8% when it does not. Intention-to-treat estimates compare parents in treatment and control schools regardless of actual participation, which is the relevant parameter for a ministry that can open a workshop but cannot compel attendance. Parents are well balanced across the two groups but treated schools are poorer than control schools. Our preferred specifications therefore adjust for individual and school covariates. The effects on participants are then recovered using an instrumental variable estimator that uses the presence of the programme in the school as an instrument for participation, and reweighting estimators (entropy balancing (Hainmueller, 2012) and covariate balancing propensity scores (Imai and Ratkovic, 2014)) that balance non-participants and participants on observed characteristics.

We find that the programme has no detectable effect on any index of French proficiency once covariates are accounted for. This null result has a natural institutional reading: most newly arrived immigrants already receive extensive language training under the CIR, and the hours OEPRE adds are modest in comparison. What the programme moves is parents' relationship to the school. Participation in school activities increases by about 25% of a standard deviation. By contrast, home-based involvement, such as helping with homework or reading with the child, does not respond, which suggests that the programme brings parents into the school more than it changes practices at home. Parents' self-esteem rises by 20% of a standard deviation, an effect consistent with the qualitative evidence on linguistic shame and regained confidence. Effects among participants are larger under both the instrumental-variable and reweighting estimators. Finally, we find no short-run effects on civic engagement, trust in institutions, sense of belonging, gender attitudes, social networks or living conditions, dimensions that sit further down the programme's theory of change.

We measure children's achievement with their most recent individual score at the national evaluations. At the intention-to-treat level, scores in French are about 10% of a standard deviation higher in treatment schools, an estimate too imprecise to be conclusive. For children whose parents participate, the picture is sharper: test scores in French increase by 25% of a standard deviation for the instrumental variable estimator and by 30% for the reweighting

estimators, with consistently no effect in mathematics. These estimates are stable when year-of-test and grade fixed effects and baseline scores are added, and their size is large for the education literature, where most interventions in high-income countries produce effects around 10% of a standard deviation (Kraft, 2020). The concentration of the effect in French, the subject closest to the programme's content and to the home language environment, fits the interpretation that children benefit from their parents' closer engagement with the school, however not directly through increased parental involvement at home.

Together, these results point to a channel of integration that operates through attachment rather than through skills. By hosting parents within a familiar institution where they already have a reason to go, the programme lowers the cost of participation and confers legitimacy on parents' involvement, and the resulting confidence and engagement benefit children directly. The predominance of women among participants also shows that school-based programmes reach immigrant mothers, a group that standard adult integration policies rarely enrol despite strong evidence of mothers' central role in child development (Lavy et al., 2022).

The chapter contributes to two strands of literature. The literature on parental involvement has established that parental investments matter for skill formation early in life (Cunha and Heckman, 2008), and the experimental evidence shows that bringing parents closer to the school improves children's behaviour and outcomes (Avvisati et al., 2010; Avvisati et al., 2014). We add causal evidence that parental attachment to the school is a self-standing driver of immigrant children's learning, distinct from the parental income and human capital channels emphasised in existing work (Foged et al., 2023). The literature on language training finds sizeable effects of instruction on immigrants' economic and civic outcomes (Heller and Slungaard Mumma, 2023; Lochmann et al., 2019; Foged et al., 2024b). Our results depart from this language-centred view as the benefits we document arise even though language proficiency does not move, and they arise for a population that language-centred programmes reach with difficulty. School attachment therefore appears as an alternative and currently underused lever of integration policy. The caveat is that it is not a substitute for everything else: school-based integration alone does not improve deeper economic and social outcomes in the short run, and the deficit in economic and socio-cultural capital that keeps home-based involvement unchanged calls for complementary instruments.

Chapter 2: Educate and Punish: Allocation, Displacement, and the Consequences of an Intermediate Juvenile Sanction

The first chapter studies how public institutions can foster engagement among populations that have limited contact with them. The second chapter turns to young people already deeply involved with the justice system and asks how the choice between alternative forms of intervention shapes their subsequent trajectories. Contact with the justice system during adolescence coincides with the years in which schooling, training and early attachment to the labour market are decided, and the existing evidence points to durable consequences: juvenile incarceration lowers high-school completion and raises adult crime (Aizer and Doyle Jr., 2015), and even pre-trial

detention reduces graduation and raises adult arrests (Baron et al., 2023). Despite the fact that international law requires that child detention be a measure of last resort (United Nations Children's Fund, 2021b), UNICEF still counted 259,000 children in detention worldwide in 2024 (United Nations Children's Fund, 2025). What justice systems do before detention, through the measures that carry most of the response to juvenile offending, remains largely undocumented, and it is this range of intermediate measures that this chapter explores.

Juvenile justice systems place a particular emphasis on education so the decision is expected to combine accountability of the offence while preserving the educational purpose of the response. Several European countries answer this tension with a residential measure combining education and constraint outside prison (United Nations Children's Fund, 2021a), and France created its own in 2002, the closed educational centre (*centre éducatif fermé*, CEF). Each year, roughly 130,000 minors receive a penal response, a few thousand are incarcerated, and about 1,000 enter a closed educational centre. The centre hosts 12 minors for a programme designed over six months, with about two adults per minor, internal schooling and vocational training, at a daily cost of 705 euros per minor in the public sector, roughly twice that of an ordinary residential placement (Cour des comptes, 2023). A centre is staffed at 26.5 full-time positions for about ten minors present, while an educator in a community setting carries 25 measures on average (Inspection générale de la justice, 2025). What separates the centre from prison, other than its resources and nature of the programme, is its legal format: the placement cannot be ordered on its own, only as an accessory to a coercive order, in 92% of cases a judicial supervision order (Direction de la Protection Judiciaire de la Jeunesse, 2021). Breaching the rules of the centre allows that order to be revoked, and revocation sends the minor to prison. The closed centre is therefore an educational programme delivered under a suspended threat of custody, conceived politically as the last step before incarceration.

Estimating what such a measure does is difficult for reasons specific to its position in the system. Minors who receive it are selected on a judicial trajectory that also predicts later outcomes, the French registers offer neither judge identifiers nor a credible instrument, and because the measure sits between community supervision and prison, its effect has no obvious direction: it must be estimated against the specific alternatives it displaces, which are themselves unobserved. I answer these limitations by using the information the decision itself rests on. Attached to each case is a social and educational report written by the duty educator before the hearing, which states the placement the educator advises the judge to order and records what the administrative trail does not: the family situation, housing, schooling, any judicial or child-protection history. Comparing minors who received the same recommendation, some placed in a centre and others not, holds constant much of what would otherwise be unobservable. I combine two administrative sources from the Ministry of Justice: PARCOURS, which links 80,201 penal measures between 2021 and 2024 to such a report, 4,302 of which carry a recommendation for a closed centre, and the historical registers GAME and IMAGES, with 934,374 spells between 2012 and 2022, where the comparison group is instead constructed by matching and reweighting minors on their recorded trajectory. The registers also carry monthly records of what happens inside each placement.

I first ask what governs the decision. The recorded judicial history predicts who will later be detained but predicts placement in a closed centre poorly. Offence characteristics contribute more to predicting detention, while the educator's recommendation contributes substantially more to predicting CEF placement. The recommendation therefore captures information relevant to the placement decision that is only weakly reflected in the administrative records. What the trajectory and the recommendation together leave unexplained seems governed less by the minor than by the receiving side of the allocation: minors are frequently placed outside their home department, facilities differ in the severity of the profiles they admit, and admissions arrive in batches, so that who lives with whom is settled by the calendar of the institution rather than by the trajectory of the minor. I then ask whom the measure actually receives. Among minors recommended for a centre but not placed, the overwhelming majority receive community supervision and fewer than one in ten detention. For most of the minors it hosts, the closed educational centre therefore operates as a residential addition to a supervision order they would have received anyway, not as a substitute for prison, and the estimates measure what happens when a minor is moved one step deeper into the system rather than what happens when a minor is spared incarceration.

Placement raises the probability of a subsequent detention episode. Within the recommended population, the probability of a detention episode within twelve months of the decision rises by 8 to 10 percentage points on a counterfactual rate of 43.5%. On the historical registers, which cover a much wider population under a weaker conditioning set, the estimate is 20.4 percentage points on a counterfactual rate of 21.6%. Estimated against each displaced alternative separately, the effect is largest against community supervision, at 21.5 percentage points, exactly where the measure is most used, and remains positive against detention, at 11 percentage points, the response the centre was created to avoid.

The timing of subsequent detention points to an important role for the legal instrument carrying the CEF placement. Among the detentions that follow within twelve months, the share that begins by ending the placement measure, which is what a revocation produces, reaches 26.8% after a closed centre against 7.1% after a supervision order alone. The monthly records add that incidents and unauthorised absences are more frequent inside closed centres than in comparable ordinary placements, and that the adverse pattern concentrates on stays interrupted before the last two-month phase that prepares the exit. The revocation device also matters for identification: it attaches to the measure and not to the minor, so a comparison minor under the same supervision order has no residential rules to breach and no facility to run from.

The design rests on selection on observables, and I run three exercises to assess the robustness of the results to unobserved selection. Holding the educator's recommendation constant moves the 12-months estimate from 17 to 14 percentage points, so the information the institution acted on and the analyst usually does not observe does not drive the result. Following Oster (2019)'s proportional-selection exercise, I find that unobserved selection would have to be 4 times as strong as selection on the observed covariates to eliminate the 12-months estimate on the historical registers sample. On the population recommended for a CEF, the coefficients moves away from zero as covariates are added. I combine this approach with Cinelli and Hazlett

(2020)'s partial-R^2 which shows that eliminating the 12-months estimate would require a single confounder explaining 7.1% of the residual variation in both placement *and* detention. On the sample of minors with an existing recommendation, such a confounder would have to be almost 3 times as predictive as the educator's recommendation on *both* placement and detention. These results are not evidence that the CEF itself is ineffective as a rehabilitation programme. The estimates capture the effect of placement relative to the alternative measure it displaces, not the effect of completing a CEF stay rather than interrupting one. Moreover, the outcomes are limited to judicial trajectories and observed before age 18, so the analysis cannot assess effects on schooling, employment and health, which are central to the centre's rehabilitative goals.

The chapter makes four contributions to the economics of juvenile justice. It treats the allocation of an intermediate sanction as an object of study rather than as a source of variation; it measures the displaced counterfactual instead of assuming it, one alternative at a time; it shows that the revocation channel documented for probation and parole (Rose, 2021; Arbour and Marchand, 2025) also operates on a measure the law presents as an alternative to incarceration; and it provides evidence on the intermediate measure of the penal ladder, whereas the juvenile literature has concentrated on the custodial margin and remains divided (Hjalmarsson, 2009; Lotti, 2022; Bhuller et al., 2020). The findings also have a direct institutional counterpart: from September 2026, public-sector CEFs become units in which the supervision order of a prosecuted minor aged 13 to 16 can no longer be revoked, while accredited associative centres keep the revocation channel (Ministère de la Justice, 2026). If the device identified here is the one that carries the effect, the two sectors should diverge, and the evaluation of the reform will offer a cleaner test than the one available today. The final chapter returns to the opposite margin of institutional reach: a population with acute needs and formal access to a service, which nevertheless makes almost no use of it.

Chapter 3: The Impact of Lifting Barriers to Mental Health Care for Refugees

Co-authored with Flore Gubert, Marie-Caroline Saglio-Yatzimirsky, Hala Tobji and Jérôme Valette

Where the previous chapter studies the consequences of intensive institutional intervention, this final chapter examines the opposite problem: how to generate effective access when services are formally available but scarcely used. Indeed, many welfare-enhancing services go unused by the very people they are designed for; take-up of social benefits is incomplete in most countries and programmes (Currie, 2004; Ko and Moffitt, 2022), and a large experimental literature shows that light-touch interventions, a reminder, a simplified form, a piece of information, can raise take-up substantially when the binding constraints are informational or administrative (Bhargava and Manoli, 2015; Domurat et al., 2021). Whether this logic extends to services whose barriers are also socio-cultural, rooted in stigma and mistrust, is far less understood. The third and final chapter studies this question in the setting of access to mental health care for refugees where

these barriers are particularly intense.

Forced displacement reached record levels in 2025, and the refugee population living in Europe has almost doubled since 2021 (UNHCR, 2025). France alone received about 145,000 asylum applications in 2025 (OFPRA, 2025). Refugees exhibit persistently lower employment rates and poorer housing conditions than other immigrant groups (Brell et al., 2020; Fasani et al., 2021), and mental health is increasingly recognised as one reason why: post-traumatic stress disorders are widespread among displaced populations and causally reduce their employment (Dang et al., 2023; Foged et al., 2026), in line with broader evidence that mental illness depresses labour market outcomes (Ridley et al., 2020). In France, 47% of newly protected refugees report mental health problems one year after obtaining their status, and the figure is still 30% four years later (ELIPA2, 2024). Refugees are nonetheless entitled to free and specialised care. Using it requires knowing that services exist, navigating an unfamiliar healthcare system in a foreign language, and overcoming a stigma that remains particularly strong in many communities. Among Syrian refugees in Jordan for instance, information about free mental health counselling spreads much more easily when a monetary incentive gives senders a social cover for sharing it (Smith, 2025). This combination of structural and socio-cultural barriers is precisely what makes refugee mental health care a demanding test for light-touch policies.

We try to answer this question using a randomised experiment embedded in AGIR, the French government's main integration programme for recognised refugees. This programme provides a 24-month individualised support pathway towards employment and housing to the most vulnerable of them, as 78% of entrants are neither employed nor in vocational training (DSED, 2026), but includes no health component. In partnership with France Terre d'Asile and COALLIA, the two organisations operating the programme in Paris and Essonne, we randomly assigned 1,403 consenting refugees to one of four groups. The control group receives a leaflet in its own language listing mental health services suited to refugees. A first treatment group has, in addition, a one-off confidential discussion with a trained peer-helper, most often a refugee with a similar background, speaking the participant's language, designed to lift stigma and mistrust towards mental health professionals. A second treatment group receives access to a multilingual helpline staffed by the same peer-helpers, who book a first appointment on the caller's behalf at a partner facility, designed to lift navigation barriers. A third treatment group receives both. Randomisation is stratified by *nationality group* × *gender*, and AGIR caseworkers do not know individual assignments. The results reported here cover the 1,135 participants of the Paris site[5].

Because asking people about a stigmatised behaviour can be unreliable, our primary outcomes are administrative traces of contact with care, collected from three sources: the helpline logs managed by our team, the records of partner mental health facilities, and admissions in the Paris public hospital network. The monitoring system of AGIR adds complete employment and housing trajectories and programme exits, a baseline survey characterises participants, and an endline survey in early 2027 will measure mental health and attitudes towards care directly.

We find that offering the two interventions together more than doubles engagement with

[5]The intervention in the Essonne department's site ended in mid-June, and is too recent to be informative. It is therefore not included in the version of this paper.

care: the probability of any recorded contact rises by 7.5 percentage points, against 6.2% in the control group. The helpline alone triggers its intended behaviour, raising the probability of a call by 3 percentage points, but moves the overall measure by an insignificant amount, and the peer discussion alone moves nothing we observe. When calls occur in the combined group, about half convert into a scheduled appointment and almost all scheduled appointments are attended. Most of the additional contact, however, does not show up at our partner facilities but takes the form of hospital admissions: the intervention seems to make care salient, and refugees then turn to the mainstream system rather than to the project's network, which matters for scalability since the response does not depend on project-specific structures. The effects are suggestively concentrated among women, for whom the combined treatment raises the probability of any recorded contact by 16 percentage points against 6 for men, although the difference between the two is not statistically significant. In absolute terms these effects are modest; in proportional terms they sit at the upper end of what light-touch interventions achieve, where reminder letters raise marketplace enrolment by 16% (Domurat et al., 2021). They are also lower bounds, since any care obtained outside our three sources leaves no trace.

At this horizon, the interventions leave employment and housing unchanged: control refugees are active on about 64% of their observed days and in stable housing on 20%, and no arm departs from these levels. Given the modest movement in care and the fact that mental health is a slow-moving input into economic integration, the follow-up may simply be too short, a question the forthcoming endline and matching with administrative data are designed to settle. One pattern resists explanation: refugees in the combined group are less likely to have already left the programme with a stable housing or employment solution recorded, although they hold stable housing on the same share of days as control refugees. We can rule out mechanical explanations but document no channel, and report it as an unexplained regularity that longer follow-up will resolve, since AGIR support is capped at 24 months and every participant eventually exits. The null effect of the peer discussion is also puzzling but it may be that a single conversation may be too small a dose against a barrier compared to similar interventions with dozens of sessions (Islam et al., 2024). Additionally, the contrast with Smith (2025), who shows that Syrian refugees share mental health information much more easily once a payment provides social cover, suggests that the stigma barrier is real but too persistent to be removed by a single conversation. Alternatively, it may also change willingness to seek care in ways that we are not able to measure before these attempts converts into real appointments. The endline survey, which will measure stigma, will allow us to separate the two, and a reconstruction of which peer-helper each participant would have faced suggests in the meantime that the message matters more than the messenger.

The chapter provides one of the first randomised evaluations of interventions against demand-side barriers to mental health care among recognised refugees in a high-income country. Existing experimental work on mental health support for displaced populations comes from humanitarian settings in low- and middle-income countries (Islam et al., 2024; Moya et al., 2024) or from undocumented immigrants in the United States (Sabety et al., 2023), while the literature on refugee integration in rich host countries has focused on language, financial assistance, legal

status or job search (Dustmann et al., 2017; Foged et al., 2024a) and paid little attention to mental health as a policy lever. Our four-arm design separates the socio-cultural and structural barriers that this literature usually treats jointly, and the intervention was built to be directly scalable, as a mental health module grafted onto an existing national programme rather than a stand-alone protocol. The results so far suggest that engagement with formal mental health care can be increased, modestly, in the arm combining interventions targeting stigma and navigation barriers.

References

Aizer, Anna and Joseph J. Doyle Jr. (2015). "Juvenile Incarceration, Human Capital, and Future Crime: Evidence from Randomly Assigned Judges". *Quarterly Journal of Economics* 130.2, pp. 759–803 (cit. on p. 6).

Alsan, Marcella, Owen Garrick, and Grant Graziani (2019). "Does Diversity Matter for Health? Experimental Evidence from Oakland". *American Economic Review* 109.12, pp. 4071–4111 (cit. on p. 1).

Alsan, Marcella and Marianne Wanamaker (2018). "Tuskegee and the Health of Black Men". *The Quarterly Journal of Economics* 133.1, pp. 407–455 (cit. on p. 1).

Arbour, William and Steeve Marchand (2025). "Can Parole Reduce Both Time Served and Crime?" (Cit. on p. 9).

Avvisati, Francesco, Bruno Besbas, and Nina Guyon (2010). "Parental Involvement in School: A Literature Review". *Revue d'économie politique* 120.5, pp. 759–778 (cit. on pp. 4, 6).

Avvisati, Francesco, Marc Gurgand, Nina Guyon, and Eric Maurin (2014). "Getting Parents Involved: A Field Experiment in Deprived Schools". *Review of Economic Studies* 81, pp. 57–83 (cit. on p. 6).

Baron, E. Jason, Brian Jacob, and Joseph Ryan (2023). "Pretrial juvenile detention". *Journal of Public Economics* 217, p. 104798 (cit. on p. 7).

Bhargava, Saurabh and Dayanand Manoli (2015). "Psychological Frictions and the Incomplete Take-Up of Social Benefits: Evidence from an IRS Field Experiment". *The American Economic Review* 105.11, pp. 3489–3529 (cit. on pp. 1, 9).

Bhuller, Manudeep, Gordon B. Dahl, Katrine V. Løken, and Magne Mogstad (2020). "Incarceration, Recidivism, and Employment". *Journal of Political Economy* 128.4, pp. 1269–1324 (cit. on p. 9).

Brell, Courtney, Christian Dustmann, and Ian Preston (2020). "The labor market integration of refugee migrants in high-income countries". *Journal of Economic Perspectives* 34.1, pp. 94–121 (cit. on p. 10).

Castell, Laura, Marc Gurgand, Clément Imbert, and Todor Tochev (2025). "Take-up of Social Benefits: Experimental Evidence from France". *American Economic Journal: Economic Policy* 17.4, pp. 1–29 (cit. on p. 2).

Cinelli, Carlos and Chad Hazlett (2020). "Making Sense of Sensitivity: Extending Omitted Variable Bias". *Journal of the Royal Statistical Society Series B* 82.1, pp. 39–67 (cit. on p. 8).

Cour des comptes (2023). *Les centres éducatifs fermés et les établissements pénitentiaires pour mineurs*. Tech. rep. S2023-0733. Cour des comptes (cit. on p. 7).

Cunha, Flavio and James J Heckman (2008). "Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation". *Journal of human resources* 43.4, pp. 738–782 (cit. on p. 6).

Currie, Janet (2004). *The take up of social benefits* (cit. on pp. 1, 9).

Dang, Hai-Anh H., Trong-Anh Trinh, and Paolo Verme (2023). "Do refugees with better mental health better integrate? Evidence from the Building a New Life in Australia longitudinal survey". *Health Economics* 32.12, pp. 2819–2835. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hec.4750 (cit. on p. 10).

Deshpande, Manasi and Yue Li (2019). "Who Is Screened Out? Application Costs and the Targeting of Disability Programs". *American Economic Journal: Economic Policy* 11.4, pp. 213–248 (cit. on p. 1).

Direction de la Protection Judiciaire de la Jeunesse (2021). *Étude sur le profil des mineurs placés en CEF au 15 juin 2021*. Rapport. Paris: Ministère de la Justice, Sous-direction des missions de protection judiciaire et d'éducation, Service de l'évaluation, de la recherche et du contrôle (cit. on p. 7).

Domurat, Richard, Isaac Menashe, and Wesley Yin (2021). "The role of behavioral frictions in health insurance marketplace enrollment and risk: Evidence from a field experiment". *American Economic Review* 111.5, pp. 1549–1574 (cit. on pp. 9, 11).

DREES (2022). *Mesurer régulièrement le non-recours au RSA et à la prime d'activité : méthode et résultats*. Les Dossiers de la DREES 92. Direction de la recherche, des études, de l'évaluation et des statistiques (cit. on p. 2).

DSED (2026). *Le programme d'accompagnement global et individualisé des réfugiés - AGIR Les entrées dans le programme* (cit. on p. 10).

Dustmann, Christian, Francesco Fasani, Tommaso Frattini, Luigi Minale, and Uta Schönberg (2017). "On the economics and politics of refugee migration". *Economic policy* 32.91, pp. 497–550 (cit. on p. 12).

ELIPA2 (2024). *Le parcours d'intégration des réfugiés* (cit. on p. 10).

Fasani, Francesco, Tommaso Frattini, and Luigi Minale (2021). "(The Struggle for) Refugee integration into the labour market: Evidence from Europe". *Journal of Economic Geography* 22.2, pp. 351–393. eprint: https://academic.oup.com/joeg/article-pdf/22/2/351/42902602/lbab011.pdf (cit. on p. 10).

Finkelstein, Amy and Matthew J. Notowidigdo (2019). "Take-Up and Targeting: Experimental Evidence from SNAP". *The Quarterly Journal of Economics* 134.3, pp. 1505–1556 (cit. on p. 1).

Foged, Mette, Linea Hasager, and Giovanni Peri (2024a). "Comparing the Effects of Policies for the Labor Market Integration of Refugees". *Journal of Labor Economics* 42.S1, S335–S377 (cit. on pp. 4, 12).

Foged, Mette, Linea Hasager, Giovanni Peri, Jacob Nielsen Arendt, and Iben Bolvig (2023). "Intergenerational spillover effects of language training for refugees". *Journal of Public Economics* 220, p. 104840 (cit. on p. 6).

— (2024b). "Language Training and Refugees' Integration". *Review of Economics and Statistics* 106.4, pp. 1157–1166 (cit. on p. 6).

Foged, Mette, Karen-Inge Karstoft, and Edith Zink (2026). *PTSD and Refugees' Underemployment: Evidence from Displaced Ukrainians*. Discussion Paper Series 019/26. ROCKWOOL Foundation Berlin (RFBerlin) (cit. on p. 10).

Guryan, Jonathan, Erik Hurst, and Melissa Kearney (2008). "Parental education and parental time with children". *Journal of Economic perspectives* 22.3, pp. 23–46 (cit. on p. 4).

Hainmueller, Jens (2012). "Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies". *Political Analysis* 20.1, pp. 25–46 (cit. on p. 5).

Heller, Blake and Kirsten Slungaard Mumma (2023). "Immigrant Integration in the United States: The Role of Adult English Language Training". *American Economic Journal: Economic Policy* 15.3, pp. 407–437 (cit. on p. 6).

Herd, Pamela and Donald Moynihan (2025). "Administrative Burdens in the Social Safety Net". *Journal of Economic Perspectives* 39.1, pp. 129–150 (cit. on p. 1).

Hjalmarsson, Randi (2009). "Juvenile Jails: A Path to the Straight and Narrow or to Hardened Criminality?" *Journal of Law and Economics* 52.4, pp. 779–809 (cit. on p. 9).

Hoover-Dempsey, Kathleen V and Howard M Sandler (1995). "Parental involvement in children's education: Why does it make a difference?" *Teachers college record* 97.2, pp. 310–331 (cit. on p. 4).

Imai, Kosuke and Marc Ratkovic (2014). "Covariate Balancing Propensity Score". *Journal of the Royal Statistical Society Series B: Statistical Methodology* 76.1, pp. 243–263. eprint: https://academic.oup.com/jrsssb/article-pdf/76/1/243/49514325/jrsssb_76_1_243.pdf (cit. on p. 5).

Inspection générale de la justice (2025). *Mission thématique sur les Centres Éducatifs Fermés : Rapport définitif*. Tech. rep. N°021 - 2024/00151. Inspection générale de la justice, Ministère de la Justice (cit. on p. 7).

Islam, Asad, Tanvir Ahmed Mozumder, Tabassum Rahman, Tanvir Shatil, and Abu Siddique (2024). *Forced Displacement, Mental Health, and Child Development: Evidence from the Rohingya Refugees*. Tech. rep. IZA Discussion Papers (cit. on p. 11).

Kim, Sungwon (2022). "Fifty years of parental involvement and achievement research: A second-order meta-analysis". *Educational Research Review* 37, p. 100463 (cit. on p. 4).

Kleven, Henrik Jacobsen and Wojciech Kopczuk (2011). "Transfer Program Complexity and the Take-Up of Social Benefits". *American Economic Journal: Economic Policy* 3.1, pp. 54–90 (cit. on p. 1).

Ko, Wonsik and Robert A Moffitt (2022). "Take-up of social benefits". *NBER Working Paper* w30148 (cit. on pp. 1, 9).

Kraft, Matthew A (2020). "Interpreting effect sizes of education interventions". *Educational Researcher* 49.4, pp. 241–253 (cit. on p. 6).

Lavy, Victor, Giulia Lotti, and Zizhong Yan (2022). "Empowering mothers and enhancing early childhood investment: Effect on adults' outcomes and children's cognitive and noncognitive skills". *Journal of Human Resources* 57.3, pp. 821–867 (cit. on p. 6).

Lochmann, Alexia, Hillel Rapoport, and Biagio Speciale (2019). "The effect of language training on immigrants' economic integration: Empirical evidence from France". *European Economic Review* 113, pp. 265–296 (cit. on p. 6).

Lotti, Giulia (2022). "Tough on Young Offenders: Harmful or Helpful?" *Journal of Human Resources* (cit. on p. 9).

Ministère de la Justice (2026). *Circulaire de politique pénale et éducative relative à la justice des mineurs*. Circulaire NOR JUSF2604225C. Paris: Ministère de la Justice, Garde des sceaux (cit. on p. 9).

Moya, Andrés, Maria José Torres, Juliana Sánchez-Ariza, Arturo Harker, Alicia Lieberman, Blasina Nino, et al. (2024). "Caregiver Mental Health and Early Childhood Development: Experimental Evidence from a Conflict-Affected Setting" (cit. on p. 11).

Nations, United (2024). *International Migrant Stock 2024: Key facts and figures*. Report 13. UN DESA/POP/2024/DC (cit. on p. 3).

OECD (2025). *Résultats du PISA 2022 (Volume I) : Apprentissage et équité dans l'éducation, PISA*. OECD (cit. on p. 4).

OFPRA (2025). *Office français de protection des réfugiés et apatrides. Rapport d'activité 2024* (cit. on p. 10).

Oster, Emily (2019). "Unobservable Selection and Coefficient Stability: Theory and Evidence". *Journal of Business & Economic Statistics* 37.2, pp. 187–204 (cit. on p. 8).

Ridley, Matthew, Gautam Rao, Frank Schilbach, and Vikram Patel (2020). "Poverty, depression, and anxiety: Causal evidence and mechanisms". *Science* 370.6522, eaay0214 (cit. on p. 10).

Rose, Evan K. (2021). "Who Gets a Second Chance? Effectiveness and Equity in Supervision of Criminal Offenders". *Quarterly Journal of Economics* 136.2, pp. 1199–1253 (cit. on p. 9).

Sabety, Adrienne, Jonathan Gruber, Jin Yung Bae, and Rishi Sood (2023). "Reducing Frictions in Health Care Access: The ActionHealthNYC Experiment for Undocumented Immigrants". *American Economic Review: Insights* 5.3, pp. 327–46 (cit. on p. 11).

Smith, Emma C. (2025). *Stigma and Social Cover: A Mental Health Care Experiment in Refugee Networks* (cit. on pp. 10, 11).

Turney, Kristin and Grace Kao (2009). "Barriers to school involvement: Are immigrant parents disadvantaged?" *The Journal of Educational Research* 102.4, pp. 257–271 (cit. on p. 4).

UNHCR (2025). *UNHCR, Global Appeal 2025*. Tech. rep. (cit. on p. 10).

United Nations Children's Fund (2021a). *Achieving Justice for Children: A Review of Innovative Data Initiatives around the World*. Tech. rep. New York: UNICEF (cit. on p. 7).

— (2021b). *Estimating the Number of Children Deprived of Liberty in the Administration of Justice*. Tech. rep. New York: UNICEF (cit. on p. 7).

— (2025). *Children Deprived of Liberty in the Administration of Justice: Global and Regional Estimates and Trends*. Tech. rep. New York: UNICEF (cit. on p. 7).

Chapter 1

Opening Schools to Parents to Promote Migrants' Inclusion and their Children's Achievement

Evidence from a large-scale French Programme

This chapter is co-authored with Flore Gubert, Élise Huillery and Jean-Noël Senne.

Abstract

The integration of immigrants is typically studied through the lenses of language acquisition and labour market outcomes. This paper highlights an alternative mechanism: attachment to host-country institutions. We study whether schools, as one of the first institutions immigrant parents interact with, can foster integration by strengthening parents' engagement with the educational environment. Using quasi-experimental variation in access to a French programme targeting allophone immigrant parents, combined with original survey data and administrative records on children's achievement on a sample of about 700 parents and 500 children, we estimate the effects of school-based parental integration on both parents and children. We find large increases in parents' participation in school activities and self-esteem, but no detectable effects on language proficiency. Children of participating parents also achieve about 0.30 standard deviations higher scores in French. The paper contributes to the literature on immigrant integration by showing that parental attachment to schools constitutes an efficient channel of integration, operating independently of language acquisition, by fostering confidence, legitimacy, and engagement among disadvantaged populations, with substantial benefits for educational outcomes of their children.

JEL codes: F22, I21, J15, J16, J24
Keywords: Immigrant integration; Parental involvement; School-based intervention; Women immigrants, Language training

We are deeply indebted to Lina El Chamaa for her insightful and determinant qualitative work, Charlène Belma and Lucie Fourel for outstanding project management, Sheyma Achir, Fanta Diakhite, Juliette Michelet, Maxime Petrosino and Isoline Rossi for their excellent supervision of the field work, as well as Lucie Tissot and Maëlle Blot for great data research assistance. We also thank all the field assistants for their discipline and enthusiasm in recruiting the parents, as well as phone interviewers for collecting the survey data. We are grateful to the staff at the French Ministry of Education (Direction de l'évaluation, de la prospective et de la performance and Direction Générale de L'Enseignement Scolaire) for their assistance in accessing and collecting the data. This study received IRB approval from the Paris School of Economics on 26 May 2023 (2023-012). Access to DEPP data was made possible by the Innovation, Data and Experiments in Education (IDEE) programme, funded by the National Research Agency (ANR) as part of the "Investments for the Future" programme whose reference is: ANR-21-ESRE-0034. This research was supported by grants from the Direction générale des étrangers en France (DGEF) under convention 2023-10-25, and BNP Paribas' Company Engagement Department under conventions 2020-10-01 and 2022-11-21.

Contents

1 Introduction . . . **19**
2 Context and Programme Description . . . **22**
2.1 Newly Arrived Immigrants' Integration Pathway in France . . . 22
2.2 The OEPRE Programme . . . 24
3 Research Design . . . **26**
3.1 Pilot and Qualitative Study . . . 26
3.2 Sampling Strategy . . . 28
3.3 Timeline . . . 31
4 Data and Descriptive Statistics . . . **31**
4.1 Data Sources . . . 31
4.2 Description of Parents and Schools . . . 32
4.3 Outcomes of Interest . . . 34
5 Estimation Strategy . . . **35**
5.1 Balance Checks between Schools and Parents . . . 35
5.2 Intention-to-Treat Estimates . . . 36
5.3 Effects on Programme Participants . . . 37
6 Results . . . **39**
6.1 Intention-to-Treat Estimates . . . 39
6.2 Effects on Programme Participants . . . 42
7 Conclusion . . . **47**
References . . . **49**
Appendix . . . **51**
A Additional Tables and Figures . . . **51**
B Construction of the Variables and Indices . . . **57**
B.1 Outcomes . . . 57
B.2 Control variables . . . 62
C Main Estimates . . . **64**

1 Introduction

Over the last three decades, the share of immigrants in Europe and Northern America has nearly doubled, rising from around 8 percent in 1990 to roughly 14 percent in 2024 (Nations, 2024). Against this backdrop of growing immigration, understanding how immigrants integrate into host societies has become an increasingly important policy challenge. Immigrant integration is often studied through the lens of language acquisition or labour market outcomes. Yet for many immigrant parents, and mothers in particular, everyday contact with host-country institutions happens first and foremost through their children's school. Schools are thus a potentially powerful, but underused, entry point for integration policy: unlike stand-alone language or employment programmes, they offer a recurring setting in which parents already have a reason to show up. Understanding whether, and through which channels, schools can serve as effective sites for immigrant integration is therefore a policy-relevant question that goes beyond the evaluation of any single programme.

Immigrant mothers are notoriously hard to reach through conventional integration programmes. Lower language proficiency, childcare responsibilities, and social norms that often confine them to domestic activities all limit their engagement with language classes or employment services located outside familiar spaces, even though comparative evidence from Denmark suggests that human capital policies such as language training are among the few that benefit immigrant women as much as men, leading Foged et al., 2024a to suggest that targeting these policies towards women may maximise their long-run impact. Additionally, legal status often compounds these barriers: many immigrant women hold residence permits obtained through family reunification and are tied to their spouse. This dependence has been shown to hamper the integration of refugee women in Denmark (Hasager, 2024). As a result, evidence on effective ways to engage this population and on the downstream consequences for maternal empowerment, family well-being, and children's schooling remains scarce, despite growing awareness that mothers have particularly strong effects on child development (Lavy et al., 2022).

The consequences of this evidence gap extend beyond immigrant mothers themselves. A large body of research shows that parental involvement in schooling is a key determinant of children's academic success (Hoover-Dempsey and Sandler, 1995; Avvisati et al., 2010; Kim, 2022), operating through the accumulation of social capital and the alignment of expectations between home and school (Hill and Taylor, 2004). But immigrant parents are consistently less involved in their children's schooling than native-born parents (Lareau, 1987; Turney and Kao, 2009; Guryan et al., 2008; Antony-Newman, 2019; Jung and Zhang, 2016), which likely contributes to the persistent academic gap observed between students of immigrant background and their native peers, even net of socio-economic background and home language (OECD, 2025; Alieva et al., 2024; Tucci et al., 2026). This naturally raises the following question: can interventions that engage parents through the school setting rather than through language training alone close some of this gap?

We study this question using the OEPRE programme (*Ouvrir l'École aux Parents pour la Réussite des Enfants*), which offers a distinctive setting to test it. Implemented and funded

by the Ministry of Education and the Ministry of the Interior, this programme is designed for foreign or immigrant parents (especially allophone, *i.e.* non-French-speaking parents) whose children are enrolled in the French school system (Ministère de l'Éducation nationale, 2025). It provides free training sessions held in schools aimed at strengthening parents' ability to support their children's schooling. Unlike stand-alone language courses, OEPRE classes are held directly within the schools attended by participants' own children, bringing together mostly mothers who share the same daily reality and are already anchored in the school through their children. Its content goes beyond language instruction to include the workings of the French school system and civic and republican values, explicitly targeting parental involvement and integration rather than language proficiency alone (Ministère de l'Éducation nationale, 2025). OEPRE therefore lets us isolate the effect of a school-anchored, community-based intervention from that of language acquisition *per se*, a distinction that most existing studies of language training programmes cannot make, since language instruction is typically their sole component.

Based on a quasi-experimental approach, our impact assessment aims to determine the extent to which the programme achieves the following three objectives: (i) involving immigrant parents in their child's schooling; (ii) promoting children's learning and success; and (iii) promoting the socio-economic integration of parents. We find that school-level exposure to the OEPRE programme significantly increases parental participation in school activities. The effect size is substantial, approximately 0.25 standard deviations (hereafter, SD), which is sizeable by the standards of the impact evaluation literature, and is twice as large among parents who effectively participate in the programme. The programme thus reaches a population that conventional integration policies struggle to enrol, as participants are predominantly mothers, recruited through the school their children already attend. Strikingly, this effect operates largely independently of language gains: language proficiency itself shows no significant improvement. Instead, we find a significant 0.20 SD increase in parental self-esteem in schools hosting the programme, suggesting that OEPRE's impact runs through a channel of school attachment and legitimacy, rather than through language skills *per se*. Most importantly, the programme significantly increases the test scores in literacy of children whose parents participate, by about 0.30 SD. This effect size is quite remarkable in the literature, since most educational interventions in high-income countries have no effect or small-sized effects of less than 0.10 SD (Kraft, 2020). Together, these results suggest that schools can serve as effective sites for immigrant parent integration, and that programmes explicitly designed to build parents' confidence and connection to the school (as opposed to language skills alone) may be a particularly promising, and currently underused, lever to boost immigrant children's educational outcomes.

This paper contributes to two strands of the literature. It first relates to the literature on the importance of parental involvement in school. Historically approached through the lens of sociology and psychology, this field has been enriched more recently by economists, aiming to provide credible identification of the causal effects of parental involvement on children's cognitive development, learning attitudes, and academic outcomes (see Avvisati et al. (2010) for a review). Most existing studies use either structural (Todd and Wolpin, 2007; Cunha and Heckman, 2008; Cunha et al., 2010), or reduced-form approaches (Aizer, 2004; Kalb and

Van Ours, 2014; Welsch and Zimmer, 2008) and generally find that parental investments are important for skill production early in life. Cunha and Heckman, 2008, for instance, estimate a dynamic factor model to study how cognitive and non-cognitive skills evolve over childhood, and how family environments and parental investments shape them at different stages of the life cycle. They find that parental inputs have different effects at different stages of the child's life cycle, with cognitive skills more affected at early ages and non-cognitive skills more affected at later ages. This impact of parental involvement among younger age groups is confirmed by the limited number of experimental studies (*e.g* Avvisati et al. (2014) and Attanasio et al. (2020)). The paper adds to this literature by showing first evidence that parental attachment to schools and increased participation in school activities is a self-standing driver of immigrant children's learning outcomes, a channel distinct from the parental human capital and income improvements emphasised in the existing literature (Foged et al., 2023).

This paper also relates to the migration and integration literature by studying how improving host-country language skills among immigrant parents influences their integration. Among the recent well-identified studies estimating the causal effects of language training, Heller and Slungaard Mumma, 2023 exploit randomised enrolment lotteries in one of the largest adult English for Speakers of Other Languages (ESOL) programme in Massachusetts to identify the effects of language training on civic engagement and labour market outcomes in the US. They find that attending adult ESOL classes more than doubles voter registration rates and increases annual earnings by $2,400 (56%). They further show that the resulting increase in tax revenue fully covers programme costs over time, yielding a 6% annual return for taxpayers. Other studies conducted in European countries exploit quasi-experimental variation from regression discontinuity designs based on test scores (Lochmann et al., 2019) or admission dates (Foged et al., 2024b; Sarvimäki and Hämäläinen, 2016) and reach quite similar conclusions. The Danish evidence is one of the most comprehensive to date: the 1999 reform that expanded language training for refugees generated permanent earnings gains, which emerged only after completion of the classes (Foged et al., 2024b); quasi-random variation in refugees' commuting time to training centres shows that hours of class translate into higher proficiency, further education, and a stronger attachment to the local communities where refugees were initially settled (Foged and Werf, 2023); and a systematic comparison of Danish integration policies concludes that language training is the policy with the largest long-run labour market effects, whereas cutting initial welfare payments produces only short-lived gains (Foged et al., 2024a). Most importantly for our purposes, Foged et al., 2023 show that the benefits of the Danish language training expansion extended to the children of treated refugees, who became more likely to complete lower secondary school and less likely to commit juvenile crime, effects consistent with both an improved home environment and the parents' higher earnings. In the French case, Lochmann et al., 2019 exploit the eligibility rule of the language training programme offered by the French Ministry of the Interior, which assigns training to immigrants whose score on an initial language test falls below a certain threshold. They find that the number of assigned hours of training significantly increases labour force participation among treated immigrants, with larger effects for individuals with higher levels of education. Interestingly, the investigation of mechanisms reveals that, beyond improved

oral language skills, participants also benefit from the positive social interactions with their classmates and teachers that language classes generate. Our findings nevertheless depart from this language-centred literature: while these studies identify language proficiency and associated social interactions as the primary channel of impact, our results point to school attachment and parental legitimacy as an alternative, self-standing channel that operates even when language gains do not materialise. Besides, our paper also provides first evidence that school-based integration programmes may not be sufficient to improve deeper economic and social outcomes such as participation in the labour market, sense of belonging, or civic and political participation.

The remainder of this paper is organised as follows. Section 2 provides background information on the immigrant integration pathway in the French context and describes the OEPRE programme in more depth. Section 3 describes the research design and sampling strategy. Section 4 presents the data sources and describes the samples of schools, parents and children, as well as the outcome variables of interest. Section 5 explains the empirical strategy and discusses its validity. Section 6 presents the results, and Section 7 concludes.

2 Context and Programme Description

The OEPRE programme is part of a broader integration pathway designed for newly arrived foreign nationals in France. Specifically, it targets non-French-speaking foreign parents whose children are enrolled in the French school system. The main objective is to bridge the sociolinguistic gap between immigrant and native parents, thereby supporting their children's academic success. Additionally, the programme leverages schools as a means to facilitate the socio-economic integration of these parents, especially women immigrants who are more frequently in charge of their children's schooling.

2.1 Newly Arrived Immigrants' Integration Pathway in France

In France, the French Office for Immigration and Integration (OFII, *Office Français de l'Immigration et de l'Intégration*) manages policies for the reception and integration of newly arrived foreign nationals. Since 2016, any foreign national lawfully admitted to French territory who is obtaining a long-term residence permit for the first time (usually valid for one year and renewable) and wishes to settle in France permanently must complete a personalised Republican Integration pathway lasting five years. Entry into this pathway involves two steps: (i) an individual interview with the OFII to assess the person's situation, skills, and needs, and (ii) a written and oral language test to evaluate his or her level of French. After completing these steps, the individual signs the Republican Integration Contract (CIR, *Contrat d'Intégration Républicaine*).

The CIR has two main objectives. First, it aims to facilitate the integration of newly arrived foreign nationals by directing them to appropriate services for employment, health, education, and more. Second, it seeks to promote their autonomy by enhancing their command of the French language and helping them understand and embrace Republican values. Under the CIR, any beneficiary of a first long-term residence permit must respect the principles and values of French society and the Republic. They are also required to attend prescribed training courses,

which include :

- A compulsory civic training course that lasts four days (a total of 24 hours) and introduces participants to the values, principles, and institutions of the Republic, secularism, and the rights and duties associated with life in France. It also provides practical guidance for everyday life, including information on housing, health, education, employment, and administrative procedures.
- A language training course that is offered to CIR signatories who are assessed as below the A2 level (according to the Common European Framework of Reference for Languages – CEFR). It includes up to 600 hours of in-person training for non-readers and non-writers, or flexible distance learning for other participants.

The CIR is valid for one year and is considered fulfilled if the required training has been attended diligently and foreign nationals have not rejected the fundamental values of French society and the Republic. Compliance with the CIR, along with other residence permit requirements, enables the issuance of a multi-year residence permit (valid for 2 to 4 years) after one year of lawful residence. In certain cases, a 10-year resident card may be granted[1].

The CIR may be extended for up to one additional year for recognised legitimate reasons, upon a proposal from the OFII. In such cases, the issuance of the multi-year residence permit is postponed until the required conditions are met. The contract may also be terminated if it is found that the foreign national, without legitimate reason, does not or no longer participates in the prescribed training or fails to meet her commitments. A negative closure of the CIR due to non-compliance prevents the issuance of a multi-year residence permit and limits the renewal of the annual residence permit to a maximum of three times.

Certain categories of foreign nationals and their family members are exempt from the CIR. This includes nationals from European Union countries, Switzerland, and other states party to the European Economic Area agreement; foreign nationals born in France who have resided continuously for at least eight years and attended school in France for at least five years after the age of 10; those who have completed at least three years of secondary education in a French institution (in France or abroad) or at least one year of higher education in France; as well as holders of short-term residence permits (such as visitors, students, trainees, medical cases, temporary workers, seasonal workers, posted workers, or talent passport holders). However, any foreign national may voluntarily request to sign the CIR to benefit from the associated training and support services. Beneficiaries of international protection (refugees or those under subsidiary protection) are also subject to the CIR, but failure to comply does not affect the issuance of their residence permit or resident card. They may be additionally referred to the

[1]The law of March 7, 2016, on the rights of foreign nationals in France reformed the reception and integration system for foreigners accessing residence in France for the first time, notably by replacing the former Reception and Integration Contract (CAI, *Contrat d'Accueil et d'Integration*) with the CIR. While the two schemes were similar in structure, the previous system had more flexible compliance requirements. The law of January 26, 2024, further strengthened the requirements regarding attendance in civic and language training. It now makes the issuance of multi-year residence permits conditional upon official certification of a French language level of A2 (previously A1), and B1 in the case of an application for a resident card, as well as success in a civic examination (with a passing rate set at 80% on a 40-question multiple-choice test).

Global and Individualised Support for Refugees programme (AGIR, *Accompagnement Global et Individualisé des Réfugiés*) in their department of residence, which provides enhanced support for social and professional integration.

On average, between 80,000 and 100,000 CIRs have been signed each year since its inception in 2016[2]. At the end of the CIR, the OFII invites participants to a final interview. This meeting serves to review the completed training and provides information about local services that can further support their ongoing integration. These local initiatives primarily focus on enhancing French language learning through sociolinguistic workshops, promoting social and professional integration, and improving access to rights. The OEPRE programme is part of this ongoing Republican integration pathway. Launched in 2017, it was designed as a complement to the CIR, offering a blend of civic and language training, specifically targeting newly arrived foreign parents who have children enrolled in the French school system as a priority group, but open to other non-French-speaking parents.

Regarding their school integration, children of newly arrived foreign nationals can be quickly enrolled in the French school system. Upon arrival, their educational needs are assessed through a placement test. Based on the results of this assessment, which is conducted in their native language along with an evaluation of their French language proficiency, pupils may be placed in an educational unit specifically designed for newcomer allophone pupils (*Unité Pédagogique pour Élèves Allophones Arrivants*, or UPE2A). These units can accommodate up to 15 pupils. As a result, approximately 60% of these children are placed in UPE2A classes, while the remainder join regular classrooms with varying levels of linguistic support: 23% receive support, and 14% do not. (DEPP-MENJ, 2020). Students in UPE2A classes are enrolled in a regular class and generally participate in full-time learning of the standard curriculum. In addition, they receive between 6 and 12 hours of support from a French-as-a-foreign-language teacher, depending on their assessed needs. UPE2A teachers often manage a very multicultural classroom, which may include students with varying levels of language proficiency. After one year, most students transition back to a regular class without additional French tutoring. Typically, parents have limited contact with the school, mainly during meetings at the beginning of the school year or when interacting with teachers throughout the year, although these meetings are not mandatory. This can create a gap between immigrant families and native parents due to language barriers and a lack of understanding of the school's functioning and requirements.

2.2 The OEPRE Programme

The OEPRE programme, funded and led by the Ministries of Education and the Interior, aims to bridge the gap between immigrant and native-born parents to enhance their involvement in their children's schooling. It specifically targets foreign, non-French-speaking parents who have children enrolled in the French school system. To qualify for the programme, parents must be non-French-speaking, legally residing in France, and have entered the country within the past five years. However, in practice, any non-French-speaking immigrant can generally participate,

[2]For more details on the Republican integration pathway for newly arrived foreign nationals, see: https://www.ofii.fr/procedure/accueil-integration/

regardless of her length of stay or residence status. Enrolment in the programme is voluntary and occurs directly at the children's school.

The OEPRE programme offers free training ranging from 60 to 120 hours, spread throughout the school year from September to June. It has two main objectives: (i) to improve children's academic achievement by encouraging parental involvement in school, and (ii) to promote the integration of immigrant parents into French society. The programme consists of three distinct components:

1. French language courses are designed to help non-French-speaking parents improve their command of the French language. These courses may lead to certification through official diplomas (such as the *Diplôme Initial de Langue Française* (DILF) or *Diplôme d'Études en Langue Française* (DELF)

2. The programme also provides an introduction to how the school system operates and outlines the expectations placed on students and parents. This aims to enhance their understanding of the educational environment and its functioning.

3. Additionally, the courses introduce the principles, customs, and values of the French Republic (including liberty, equality, fraternity, and secularism) to facilitate integration into French society.

Training sessions typically occur during the day or after school hours in primary and lower secondary schools. The organization of these sessions varies by school and generally consists of two-hour time slots, scheduled once or twice a week. To create optimal learning conditions, the sessions are conducted in small groups of 8 to 15 participants and usually run for an average of 23 weeks. When the OEPRE programme is not hosted at the primary school where their children are enrolled, parents are invited to attend sessions organised at the nearest lower secondary or primary school hosting the programme. Trainers are mainly teachers of French as a foreign language (FLE) or social workers, sometimes with support from other school staff or local partners, such as libraries, employment agencies, or town halls. While some teachers have received specific training to lead workshops, most rely on peer learning or online training resources. It is important to note that schools and trainers have some flexibility at the local level to adapt the programme to meet the needs of their target populations. The qualitative study we conducted in 2022 across 14 schools in six French regions (see Section 3.1) revealed that, although the programme is generally structured around the three components mentioned above, there is notable variability in its implementation and content. This variability largely depends on the diversity of trainers and their specific skills.

Launched in 2008 as a pilot project in around thirty schools, the OEPRE programme was designed to address the intersection of education and migration policies. It was not created in isolation but was aligned with the objectives of the CIR. Initially, the programme aimed to help overcome the language barriers migrant parents face when supporting their children's education. The success of its pilot phase, particularly in increasing parental participation in school meetings and reducing absenteeism, led to its continuation and gradual expansion, with a new focus on

enhancing school-family relationships. Over time, the programme increasingly emphasised the importance of school-related parenting and explicitly connected parental involvement to children's academic success. Additionally, schools began to be viewed more broadly as a means for integrating non-French-speaking immigrants, especially women from family-based immigration, who constituted the majority of participants.

Since the pilot phase, the number of schools hosting the programme has steadily increased to over 500 across France. The number of enrolled parents rose from around 14,000 in the 2019–2020 school year to nearly 19,000 in the 2022–2023 school year, according to the latest data from the Ministry of Interior. The data also show a growing number of parents willing to participate, including non-newly arrived parents (or those who have been in France for more than five years), who now account for nearly half of the participants. Approximately 90% of participants are women, most of whom are between 30 and 50 years old[3].

Therefore, the current structure of the OEPRE programme aims to improve French language proficiency while helping participating parents gain a better understanding of the French education system and the values of the French Republic. These three components are designed, in the short term, to encourage immigrant parents to be more involved in their children's education. In the longer term, the objectives of this linguistic, educational, and civic integration are twofold. First, enhancing parental support is expected to improve children's academic outcomes. Studies show that children who receive regular parental support and grow up in households where the language of instruction is spoken tend to achieve better academic results, especially when close parental involvement bolsters their motivation and encourages perseverance in school. Additionally, linguistic and civic integration through the school system may facilitate parents' broader integration into French society.

3 Research Design

This section presents the research design of the study. We first describe the pilot and qualitative study which inspired the design, then the sampling strategy used to select the schools and parents, and finally the timeline of the study.

3.1 Pilot and Qualitative Study

At the beginning of this research, it was important to understand what drove the eligible parents' participation in the OEPRE programme, the nature of the programme, and its potential benefits in order to develop an appropriate research design. We thus conducted a pilot study in the school year 2020-2021, as well as a qualitative study in Spring 2022.

The latter was conducted in 14 primary schools across six regions (Marseille, Grenoble, Lyon, Montpellier, Paris, and Versailles). It combined ethnographic observations of the programme implementation with semi-structured interviews, both individual and in group, conducted with

[3]For further details and data on the OEPRE programme, see: https://www.immigration.interieur.gouv.fr/lintegration-des-etrangers/dispositifs-de-lintegration-en-france/ouvrir-lecole-aux-parents-pour-reussite-des-enfants

37 participating parents and several programme facilitators. Consistent with the objectives of OEPRE, the parents interviewed were predominantly women, living with a partner, mostly originating from North Africa and Sub-Saharan Africa. They generally had low levels of formal education, were either inactive in the labour market or engaged in precarious employment, and had on average three dependent children.

The interviews highlighted several salient features of the programme. First, participants largely perceive OEPRE as a means of improving their proficiency in French: acquiring basic communication skills for beginners, refining oral and written expression for more advanced participants. Parents aim to dismantle what some of them describe as a "linguistic shame", which hampers their capacity to carry out everyday tasks and navigate administrative procedures. Second, parents want to improve their interactions with school staff and ensure closer monitoring of their children's schooling, with the hope of fostering upward social mobility. Parents aim to gain confidence in their exchanges with teachers and a better understanding of school norms without relying on a partner or a relative. Thanks to the programme, they report feeling more legitimate and better integrated into the educational community. Finally, parents see OEPRE as a way to foster their social integration by creating a space for exchange and solidarity among participants. In their view, OEPRE fosters relationships that extend beyond family or community networks, thereby reducing social isolation, which is particularly pronounced among immigrant women. While parents make relatively limited reference to the component related to "learning the values of the Republic" and often struggle to define its scope, some of them nonetheless express interest in this dimension. The learning of social norms (such as forms of address and language registers) is frequently mentioned in this context. This qualitative evidence was used to design the questionnaire and indicators of impact used in the quantitative analysis presented below.

Besides, the pilot study was meant to understand how parents decided to participate in the programme and where. In the school year 2020-2021, we sent a questionnaire to 1,524 primary schools that hosted the OEPRE programme or offered eligible parents from their school to participate to OEPRE in middle school or in another primary school nearby. Using the 470 responses received from these primary schools, we conducted an analysis of the key determinants of participation in OEPRE. We found that the main determinant of participation was the fact that the programme was hosted directly within the school attended by the children of eligible parents. More precisely, this analysis showed that participation rate increased by 34pp (on a basis of 21%) when OEPRE was offered within the school, compared with the situation in which the programme took place in middle school or in another primary school in the same municipality. As a consequence, participation rates remained low in schools with many allophone migrant parents but which did not host the OEPRE programme. Since the ministry and regional authorities have limited resources, OEPRE is not offered in all schools with a large number of allophone migrant parents.Our research design therefore relies on comparing allophone parents across similar schools in terms of observable characteristics, some of which hosted OEPRE while their children were enrolled (treatment parents) and others that did not (control parents). The challenge was thus to identify schools with a large number of allophone parents, and the years,

if any, in which they hosted OEPRE. The main challenge was thus to identify schools with a large number of allophone parents, and the years, if any, in which they hosted OEPRE.

3.2 Sampling Strategy

The sub-section describes the sampling strategy used to constitute our final sample of 151 schools and 713 allophone parents. Figure 1.1 provides a graphical presentation of the sampling strategy.

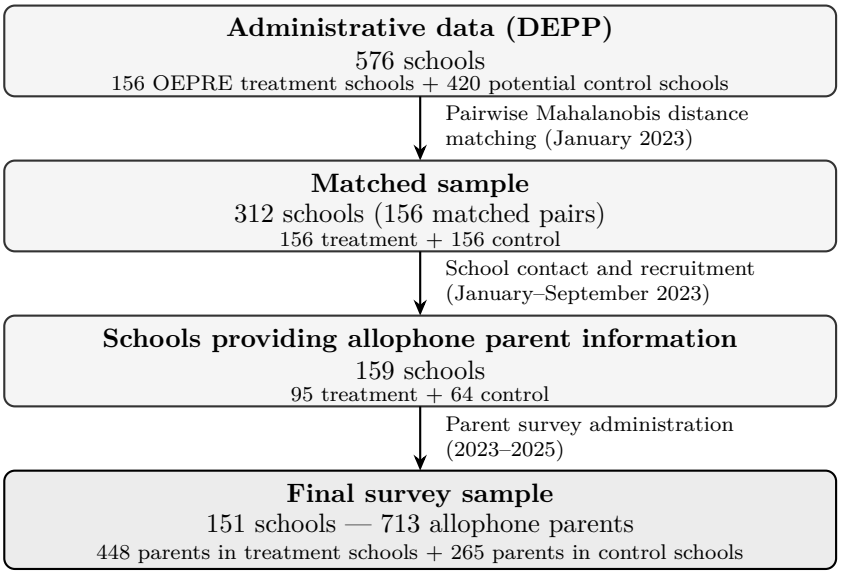

Figure 1.1: Sample Construction

Notes: Schools were contacted between January and September 2023 to obtain allophone parent lists and child enrolment information. Numbers at the school-recruitment stage (*c.* 159) are approximate. Matching uses pairwise Mahalanobis distance minimised on mean 1st-grade assessment scores (2018–2022), with exact matching on UPE2A, QPV, priority education status, EANA membership, and municipality density. Full balance statistics are reported in Table A2.

Selection of Schools With the help of regional coordinators of the OEPRE programme, an initial list of 576 primary schools was compiled, including 156 schools with OEPRE in 2022–2023, and 420 schools that they considered relatively similar but without OEPRE in 2022–2023. Note that the schools which did not host OEPRE in 2022–2023 could have hosted the programme in precedent or following years, so the treatment is not defined at the school level but at the school-year level. Nevertheless, with this list of schools, we conducted a statistical matching procedure using administrative data from the statistical department of the Ministry of Education—the *Division de l'évaluation, de la performance et de la prospective* (hereafter, DEPP). We matched each school hosting OEPRE in 2022–2023 to a school not hosting the programme the same year, by minimizing the Mahalanobis distance over a set of covariates influencing the presence or absence of the programme: being in priority education, being located in a priority urban area, the presence of a Pedagogical Unit for Allophone Students (UPE2A), the number of pupils in the school, the number of pupils per class, the number and proportion of allophone students in the school, local population density, median neighbourhood income, municipal population size, as well as average test scores at national evaluations in French and mathematics administered at

the beginning of first grade since 2018. This procedure resulted in a final sample of 312 schools in 156 pairs comprising one school with OEPRE and one school without OEPRE in 2022–2023.

We then contacted the educational authorities to obtain the institutional approval to carry out our study in these 312 schools. This process followed a clearly defined hierarchical chain, beginning with regional level (*Recteur d'académie*), followed by the district level (*Directeur académique*), then the locality level (*Inspecteur de l'éducation nationale*), and finally the school principals. We obtained the approval of all four hierarchical levels for 159 schools, among which 95 hosting OEPRE in 2022–2023 and 64 schools not hosting OEPRE in 2022–2023. In these 159 schools we then collected precise information on the years when they hosted the OEPRE programme from 2018 on to 2024, and tried to recruit allophone parents.

Selection of parents The second step consisted in contacting the principals of these 159 schools to determine the number of allophone parents with children enrolled, as well as the languages spoken by these parents. Based on this information, consent forms translated into the appropriate languages were sent to the schools, which were responsible for distributing them to the parents.

The level of engagement among school principals varied considerably. Some collaborated actively while others, facing time constraints or administrative overload, invested limited time and effort. Given the central role of school principals and teachers in distributing consent forms and raising awareness among families about the study, parents' response rates varied widely from one school to another. To increase the number of parents in the study, the research team organised numerous information meetings in schools attended by allophone parents identified and invited by the teaching staff. This allowed us to collect 865 consent forms from 151 schools.

Finally, once consent had been obtained, parents were contacted by phone to complete a questionnaire administered by interviewers speaking their preferred language. Out of the 865 consent forms collected, 713 resulted in an individual interview. The final parent and school sample therefore comprises 713 individuals from 151 distinct schools. Among these 713 parents, 448 had a child attending a school where OEPRE was offered and constitute our treatment group, while 265 had no child attending a school where the programme was offered and thus constitute our control group. As the programme could be offered in a school in some years but not other years, parents with children attending the school when it offered the programme are considered as part of the treatment group, whereas other parents with children attending the same school in other years when it did not offer the programme are considered as part of the control group. The treatment definition is thus at the parent level and not at the school level. The geographic distribution of the sampled schools is shown in Figure 1.2, with red dots representing schools offering OEPRE in 2022–2023 and blue dots representing schools not offering OEPRE in 2022–2023—similar maps for school years 2019-2020 to 2023-2024 are presented in Figure A1.

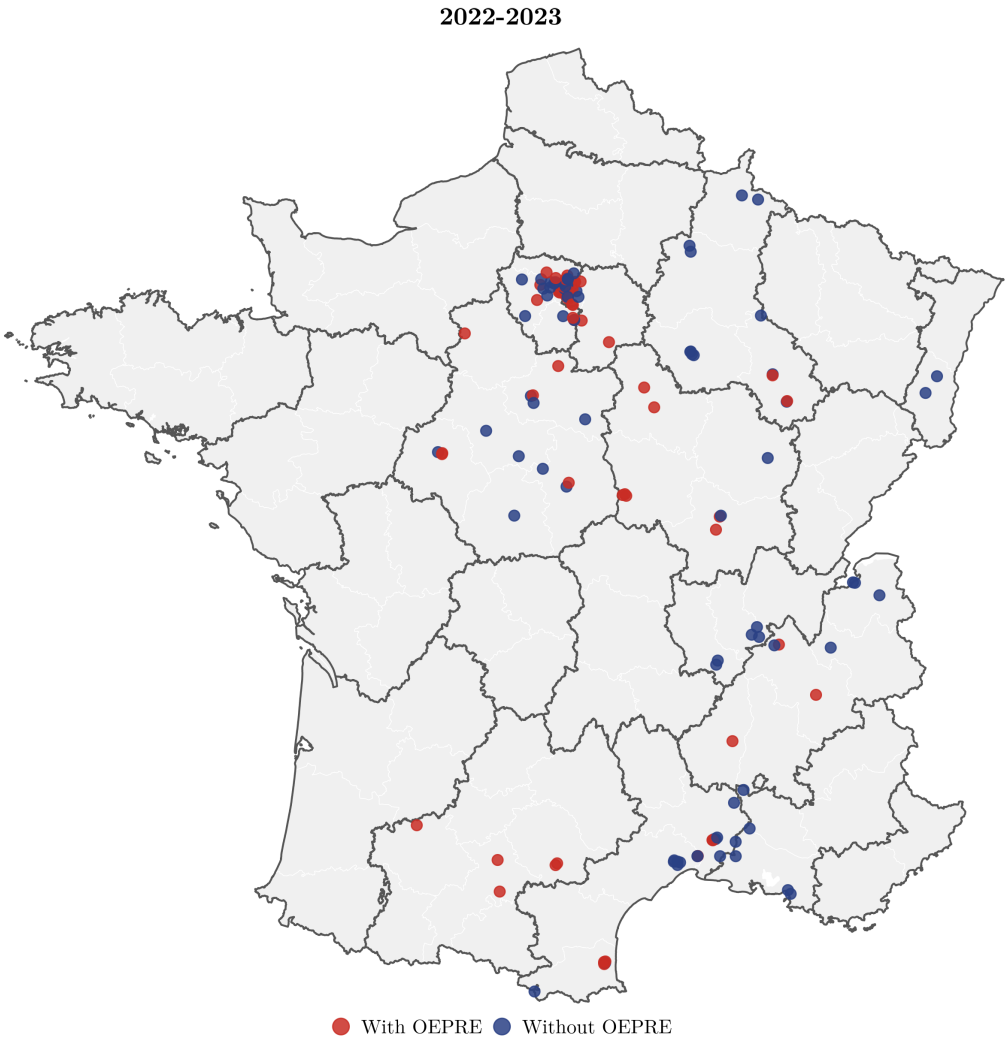

Figure 1.2: Geographic Distribution of Study Schools, 2022–2023

Notes: The map shows the 148 schools in our sample for the 2022–2023 school year: 68 schools hosting an OEPRE workshop (red dots, *With OEPRE*) and 80 schools without (blue dots, *Without OEPRE*). Schools with OEPRE status vary across years as workshops open or close. Schools are concentrated in Île-de-France and other major urban centres. See Figure A1 for the full panel covering all years from 2019–2020 to 2023–2024.

3.3 Timeline

To summarise the study design described above, Figure 1.3 provides a visual presentation of the study timeline. Overall, the study required five years of work: the pilot study in 2020-21, the qualitative study and communication with educational authorities in 2021-22, schools selection and communication with schools in 2022-23, and then field work to collect parental consents and administer parental questionnaire in 2023-24 and 2024-25.

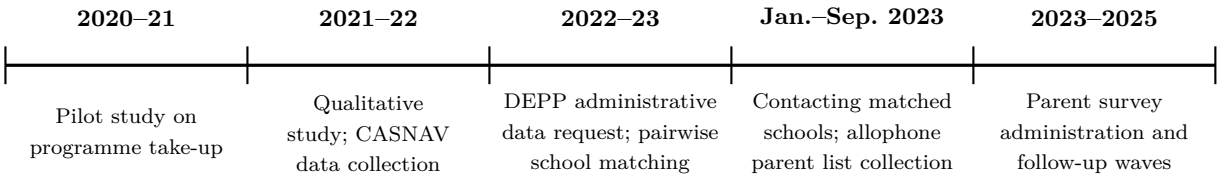

Figure 1.3: Study Timeline

4 Data and Descriptive Statistics

In this section, we describe the two sources of data used for the empirical analysis: administrative data from the Ministry of Education, and survey data from the parents.

4.1 Data Sources

Administrative data. The administrative data was provided by the statistical department of the Ministry of Education, the DEPP. They provided us with three types of information: the characteristics of schools, those of the neighbourhood, and individual students' test scores at national evaluations.

At the school level, administrative data indicates whether the school is in priority education (REP and REP+) as well as the presence of special need units to provide support to allophone students (UPE2A) and handicapped students (ULIS). The data also includes the number of students in each grade, the number of classes, and the share of girls among pupils. We also have information on teachers: average age and seniority in the school, the number of teachers, and their gender and status of teachers (tenured or not). Moreover, we observe the socioeconomic composition of the school: the average socioeconomic index of the school (IPS) constructed by the DEPP based on parents' occupation, and the shares of pupils in four socioeconomic categories defined by the DEPP which is based on parents' occupation: highly advantaged, advantaged, intermediate, and disadvantaged. Finally, we observe the average academic performance of students in the school at the national evaluations at the beginning of Grade 1, Grade 2, and Grade 4. More precisely, the data provides the share of students in three categories "At risk", "Fragile", and "Satisfactory" in both maths and French at these national evaluations.

Second, at the neighbourhood level, the administrative data indicates whether the neighbourhood is a priority urban area (QPV), which indicates that the median income of the area is below the national average income and therefore receive specific subsidies from the French

government. We also know whether the neighbourhood is located in a rural, intermediate, or dense area, the population and the median income of the municipality.

Third, the administrative data provides the individual test scores at the national evaluations. Pupils in our sample have taken different evaluations at different levels depending on their age and grade. We use 2024 and 2025 national evaluations to measure academic achievement of sampled children. We also use pre-2023 national evaluations as baseline test scores, but only for a small number of children. In fact, for most children, we do not observe pre-treatment test scores. This may be because they arrived in France too recently to have an assessment history, as recently arrived pupils do not participate in national evaluations until they have sufficient literacy skills in French. Alternatively, for children whose appropriate pre-treatment score would be the 2022 one, this is because 2022 test scores are missing from the DEPP database altogether.

Survey data. To collect individual information on parents, we conducted a survey over the phone with parents who had given their informed consent. The survey was administered in parents' preferred language, either French or their native language, by research assistants recruited and trained by our research team. The survey lasted for about an hour and asked questions on both the respondent and his/her spouse, if any.

First, the survey provides information on both parents' basic demographics such as age and sex, the structure of the family (number of children and whether the parents live together or not), and their immigration history: whether parents were raised in France, their year of arrival in France, and their country of origin. Moreover, for the respondent parent, the survey also captures nationality, language score at arrival, level of education (no education, primary, lower secondary, upper secondary, or higher education), and residency status: stable residency for those who obtained French citizenship or a residency permit for at least 10 years; temporary permit for those with a residency permit of less than 10 years residency permit; and precarious status for all other cases. All these characteristics are used as baseline variables in our analyses.

Second, the survey also provides information on outcomes potentially affected by the OEPRE programme. We describe these outcomes in Section 4.3.Before doing so, however, we present some descriptive statistics on the schools and parents included in the study.

4.2 Description of Parents and Schools

Table A1 presents the baseline characteristics of all parents in our sample and of those who, among them, participated in the OEPRE programme.

Full sample. Only 5% of respondents are French, they are 39-year-old on average, 75% are women, and their score of French proficiency was markedly low at arrival in France: 1.6 on a scale from 1 (not a all) to 4 (very well). The family organization is most often the two parents living together (79%), then the mother is alone with her children (17%), or other arrangements (4%). The family size is large since there are on average 2.7 children, two thirds of whom were born outside France. Almost half of the respondents declare that they are in the precarious

status (48%), while 22% have a temporary resident permit. Only 30% have a stable resident permit (*i.e.*, permanent or for 10 years or more).

There are a few differences between fathers and mothers. Fathers are slightly older than mothers (43- *versus* 38-year-old); 7% of them were raised in France whereas only 3% of mothers; and they arrived in France in late 2014 on average, while mothers arrived three years later in late 2017. In contrast, fathers and mothers show quite similar levels of education, with about a third having attained higher education, a third upper secondary education, and a third lower secondary education or less.

In terms of geographic origin, the sample comes from a variety of regions. The largest share comes from North Africa (29% of both fathers and mothers), then from South/East Asia (18% of fathers and 17% of mothers), then from Central/Western Asia (16% of both fathers and mothers). About 12% come from Eastern Europe, and another 8% from Western Europe/USA. Finally, less than 10% come from Sub-Saharan Africa and from Latin America.

OEPRE participants. Participation is not mandatory for allophone parents. In most cases only one parent per household participates in the workshop, and this parent is most often the mother in nearly three-quarters of cases. It is interesting to compare the characteristics of parents who volunteered to participate in the OEPRE programme with those who did not. To this end, Table A1 presents the same descriptive statistics as above for OEPRE participants, and test whether the differences with the full sample are statistically significant.

Most characteristics of participants do not differ from those of non-participants. However, there are a few significant differences. First, participants tend to have a smaller proportion of children born outside France (62% versus 67%). Second, although the vast majority of households consist of both parents, household structure seems to influence participation in the program: compared with the full sample, participating households are more often composed of both parents (86% versus 80%), and correspondingly less often consist of single-mother or other family arrangements. This difference suggests that single parenthood limits mothers' participation in the programme. The score measuring French language proficiency upon arrival in France for the responding parent is statistically identical between participants and the full sample, but it is smaller among participants with a p-value of 0.12, which suggests that a lower level of French proficiency may encourage participation in the OEPRE programme. Finally, participants are slightly less likely to come from Eastern Europe than the full sample (9% of mothers among participants compared to 13% among the full sample).

Schools. The OEPRE programme is implemented in primary schools which concentrate allophone parents, so the characteristics of schools hosting the programme are likely to differ from the average school in France. Column "Treatment" of Table A2 presents descriptive statistics on these schools, which provides evidence that they are socio-economically more disadvantaged than the average French school. One striking difference is the proportion of schools in priority education: 77% of treatment schools, whereas only 16% of schools nationally.[4] The average

[4] https://rers.depp.education.fr/

social position score of our treatment schools is 82 compared to 105 nationally.[5] Similarly, the proportion of sampled treatment schools in priority urban areas (QPV) is 74%, whereas only 16% at the national level.[6] Likewise, the average academic performance in the treatment schools lies well below the national average: the proportions of pupils in the at-risk categories in French and maths are about twice as big as the national average.[7]. All these figures underline that, unsurprisingly, our sampled treatment schools are much more disadvantaged than the typical French school.

4.3 Outcomes of Interest

Several outcome indicators were constructed from parent survey data to measure the impact of the OEPRE programme. All outcome indicators are summary indices constructed from several individual questions or items. The items were first standardised using the mean and standard deviation of parents in control schools, and then these z-scores were averaged and in turn standardised using again the mean and standard deviation of parents in control schools. The detailed items and values used to construct these summary indices are described in B.2. Outcomes are organised into two groups. The first consists of outcomes directly targeted by the OEPRE programme, such as French proficiency and school engagement. The second consists of longer-term outcomes that are not directly activated by the programme but correspond to its final objectives in the theory of change and may evolve as downstream effects.

Outcomes directly linked to OEPRE: French proficiency and engagement in schooling. The outcomes directly linked to the content of OEPRE are French proficiency and engagement in schooling. We construct three indices of French proficiency: father's level, mother's level, and linguistic fluency of the respondent. The first two indices capture mastery of French in writing, reading, speaking and understanding, whereas the third index captures respondent's ease with French language in daily and administrative life (see B.2 for more details). Besides, we construct three indices of engagement in schooling: an index of participation in school events such as back-to-school meetings with teachers, school trips, or report card meetings; an index of home activities with children such as reading a book, drawing, or cooking; and an index of involvement in schooling measuring how much parents talk with their child about school and provide support with homework.

Downstream effects on political, social, and economic life. We test the hypotheses that the programme may affect seven outcomes: civic engagement (e.g., voting or striking), trust in institutions (including the government, the education system or the press, among others), sense of belonging in France (e.g., feeling at home in France), gender attitudes measuring the extent to which respondents support gender equality, self-esteem as measured by the Rosenberg Self-Esteem Scale, social network capturing how much respondents have social relationships

[5] https://www.data.gouv.fr/datasets/indices-de-position-sociale-dans-les-ecoles-a-partir-de-2022
[6] https://www.insee.fr/fr/statistiques/8186144
[7] https://rers.depp.education.fr/

with family or friends, and finally living conditions capturing housing comfort, employment status, type of contract, hours worked weekly, and ISEI. Again, detailed questions and index construction are provided in B.2.

Academic achievement of children. We use the most recent test score obtained by each child as outcome to capture the academic achievement of children. As pupils are now assessed every year, the most recent score is from September 2025 for the vast majority of children, and from September 2024 for the remaining pupils who did not take the 2025 evaluation. In all cases, the most recent score is measured post-treatment. As noted above, for most children, we do not observe pre-treatment test scores because their assessment history is too recent and because 2022 test scores are missing from the DEPP database altogether.

5 Estimation Strategy

This section presents the balance checks between control and treatment units as well as the resulting estimations strategies used to estimate the treatment effects of the OEPRE programme.

5.1 Balance Checks between Schools and Parents

We perform balance checks to compare observable characteristics of treatment and control schools in Table A2 and of parents in treatment and control schools in Table A3. Each baseline covariate is standardised so that the mean of the control group is 0 and its standard deviation is 1. Thus, the difference between parents in the treatment group and those in the control group can be interpreted as the deviation of the treatment group relative to the control group mean, expressed in standard deviation units of the control group distribution.

The results in Table A2 show that the two groups of schools have similar characteristics on many dimensions such as the presence of language support unit (UPE2A), the number of classes and pupils, teacher characteristics, the density of the municipality, and the academic performance of students at the national evaluations. However, they differ in a few important dimensions related to one another: the treatment schools are more likely to be located in a priority urban area (QPV) (74% in the treatment group *versus* 53% in the control group), which means that the neighbourhoods of treatment schools are poorer than those of control schools. Relatedly, treatment schools are more likely to be priority education (77% *versus* 58% in the control group), their social position score is 6 points below those of control schools (82 *versus* 88), and their shares of highly advantaged and advantaged pupils are lower (9 and 7% *versus* 13 and 9%, respectively). Moreover, treatment schools are located in larger municipalities than control schools (about 288,000 inhabitants *versus* 99,000 inhabitants).

Despite these differences between treatment and control schools, the results in Table A3 show a good level of comparability between parents in the two groups: out of 42 tests, we find three statistically significant differences in means: the number of children (2.6 among control families *versus* 2.8 among treatment families, p-value = 0.09), the probability that the youngest child was born outside France (76% in the control families *versus* 62% in treatment families, p-value

= 0.02), and the probability that the father was raised in France (5% in control families *versus* 8% in treatment families, p-value = 0.06). These differences are relatively small in size. Since we would expect 10% of tests to show statistically significant differences between the two groups purely by chance under the assumption that both groups were drawn from the same population, we cannot reject the hypothesis that the two groups of parents are comparable.

All in all, the balance checks suggest that parents are quite comparable based on observable covariates, but that schools and neighbourhoods of treatment parents are more disadvantaged than those of control parents. It suggests that treatment parents are negatively selected, which we take into account in the estimation strategy presented below.

5.2 Intention-to-Treat Estimates

Identifying the causal effect of the OEPRE programme from the collected data is not straightforward due to the presence of non-compliance with treatment. Indeed, in our sample, not all parents in treatment schools participate in OEPRE even though they are, in principle, part of the programme's target population. Conversely, some parents in control schools may participate, for instance by attending OEPRE in a nearby school that offers the programme. However, participation rates in the OEPRE programme are substantially higher in treatment schools that host the programme than in control schools. Table 1.1 shows that the participation rate in the programme is only 8% among parents in control schools, while it is 54% among parents in treatment schools.

The presence or absence of a workshop in otherwise comparable schools thus induces an exogenous variation in programme participation, allowing us to estimate the causal effect of the OEPRE programme. In this context, we first produce intention-to-treat (ITT) estimates of OEPRE impact, which consists of comparing outcome indicators for parents in treatment schools with those of parents in control schools, regardless of their actual participation in the workshops. This method provides an unbiased estimate of the impact of offering OEPRE in a school, under the assumption that parents would have had the same outcomes in absence of the programme. It is particularly appropriate in contexts of partial, non-mandatory, participation among the targeted parents. It constitutes a relevant measure of the programme's effectiveness from the perspective of policymakers who can decide whether or not to open a workshop based on its effects on targeted parents and their children, without being able to compel participation.

Table 1.1: First Stage: Effect of School Assignment on Workshop Participation

	OEPRE participation
School treatment assignment	0.46***
	(0.04)
First-stage F-stat	117.40
Observations	663
R^2	0.34
Adj. R^2	0.28
Control mean	0.08
Covariates	Yes

Notes: This table reports the first-stage regression of OEPRE workshop participation on school-level treatment assignment. The dependent variable is an indicator equal to 1 if the parent attended at least one OEPRE workshop. Controls included in all specifications: father's and mother's presence in France (born/raised), year of arrival in France (father ref: before 2010; mother ref: before 2014), age group (father ref: over 48; mother ref: over 42), region of birth (8 categories; ref: Western Europe/USA), respondent's education level (ref: none), French nationality, number of children (ref: 1 child), family structure (ref: parents together), residency status (ref: stable), language score at arrival (standardised), respondent gender (mother/father). The F-statistic tests instrument relevance (Stock-Yogo weak instrument threshold: 10). $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Standard errors are clustered at the school level.

The ITT estimates are obtained by estimating the following equation with OLS:

$$Y_{i,s} = \alpha + \beta \times T_s + \boldsymbol{X}'_{i,s}\boldsymbol{\gamma} + \epsilon_{i,s} \tag{1.1}$$

where $Y_{i,s}$ is the outcome indicator of parent/child i in school s, T_s is equal to 1 if school s host the OEPRE programme, $\boldsymbol{X}_{i,s}$ are individual- and school-level baseline covariates, and $\epsilon_{i,s}$ are errors terms clustered at the school level. We run three versions of equation 1.1: without any baseline covariates, with the set of all baseline covariates—except those whose Variance Inflation Factor (VIF) is above 5, and a set of covariates selected by the double LASSO procedure. The full list of covariates considered in the second and third specifications is two-fold: the individual covariates listed in Table A1, and the school covariates listed in Table A2.

5.3 Effects on Programme Participants

In addition to ITT estimates, which capture the policy-relevant effect of offering OEPRE in a school, we also recover the effect of programme participation on parents who actually take it up. Because participation is voluntary and partly determined by parents' characteristics, we use both an instrumental variable strategy and a strategy that adjusts for selection on observables to recover, respectively, the Local Average Treatment Effect (LATE), and the Average Treatment

effect on the Treated (ATT).

Instrumental Variable Estimates. The estimation of the LATE relies on an instrumental variable strategy by which participation in OEPRE $P_{i,s}$ is instrumented by the presence of OEPRE in the school T_s (first stage) to assess its impact on outcomes $Y_{i,s}$ (second stage). The LATE estimates β' are obtained by estimating the following equations with 2SLS:

$$P_{i,s} = \alpha + \pi \times T_s + \boldsymbol{X}'_{i,s}\boldsymbol{\gamma} + \epsilon_{i,s} \tag{1.2}$$

$$Y_{i,s} = \alpha' + \beta' \times \widehat{P}_{i,s} + \boldsymbol{X}'_{i,s}\boldsymbol{\gamma}' + \varepsilon_{i,s} \tag{1.3}$$

The vector $\boldsymbol{X}_{i,s}$ in these two equations corresponds to the set of all baseline individual and school covariates, except those whose VIF is above 5. Table 1.1 above shows the results of equation 1.2: as the presence of OEPRE in a school increases the participation rate by 46pp, it is expected that the LATE estimates β' from equation 1.3 are approximately twice as large as the ITT estimates β from equation 1.1. β' provides the effect on compliers, *i.e.*, parents in treatment schools who participated in OEPRE because the programme was offered in their children's school. The first-stage F-statistic equals 117, so the instrument is highly relevant. The validity of the estimation also requires the exclusion restriction assumption to hold: treatment assignment affects outcomes only through individual participation. In the context of OEPRE, this implies that the programme has only direct effects on participants and no spillover effects on non-participating parents in treatment schools. Relatedly, our instrumental-variable design requires SUTVA, implying the absence of interference across individuals: one individual's treatment induced by the instrument must not affect another individual's potential outcomes. In particular, this requires no within-school externality between participants and non-participants.

Weighting-based Estimates. To test the robustness of the IV estimates, we use an alternative estimation strategy to recover the ATT, consisting in weighting observations in the control schools so parents' observed characteristics look similar to those of parents who participate in the programme in treatment schools. This method has an advantage over the LATE: it does not require an absence of externalities within schools, whereas this is a necessary assumption for the LATE to be valid.

To compute the ATT, we use two weighting methods: the entropy balancing method (Hainmueller, 2012) and the CBPS weighting method (Imai and Ratkovic, 2014). Both methods create weights for control units so that, after weighting, the control group looks like the treated group in terms of individual covariates. They only slightly differ in how they achieve balance and what they optimise. Entropy balancing puts a weight of 1 on treated units and optimised weight on control units so that the weighted mean of each covariate in controls equals the treated mean, and among all weights that achieve this, picks the ones that are as close as possible to uniform weights. Besides, CBPS weighting starts from the propensity score framework, but instead of estimating the propensity score solely to predict treatment, it builds balance conditions directly

into the estimation of the score. In CBPS weighting, balance is achieved in expectation but not imposed exactly as in entropy balancing. Again, treated units are weighted 1 and controls are reweighted by the propensity score.

In both methods, the set of covariates is the set of all baseline individual covariates, except those whose VIF is above 5. We do not include school covariates in the reweighing procedure because estimators fail to converge, probably because the number of observations is too small to use a large number of covariates. As a consequence, as shown in Appendix Tables A2 and A3, individual characteristics are perfectly balanced after the reweighing procedure (both CBPS and entropy), but not the school characteristics. Therefore, we include school characteristics as control variables in the ATT estimations.

6 Results

This section presents the ITT and ATT estimates of the effects of the OEPRE programme, separately for the outcomes directly targeted by the programme, for the downstream political, social and economic effects and for children's academic learning.

6.1 Intention-to-Treat Estimates

Parental Language Proficiency and Engagement in Schooling. Figure 1.4 presents the intention-to-treat estimated effects of the OEPRE programme on the primary outcome indicators specifically targeted by the programme: parental French proficiency and engagement in schooling. The precise estimates are also reported in Table C1. We favour estimates with covariates since they neutralise the differences in schools and parents observed in the previous section. The estimates without covariates are reported for comparison as a way to assess the selection bias.

The programme has no clear effect on any indicator of French proficiency: all estimates are small and insignificant, except for fathers' French level without covariates. But this difference likely reflects the fact that fathers in treatment schools are slightly more likely to have been raised in France as observed in Table A3. Once covariates are included, there is a clear absence of effect of OEPRE on language proficiency. This is not very surprising given that most immigrants benefit from the CIR which already provides many hours of French language courses as described above in Section 2.1. Our finding shows that the additional French courses provided by OEPRE do not further help immigrants improve their language proficiency, probably because the amount of hours is relatively small compared to the CIR.

However, the programme substantially increases parents' participation in school activities. The effect on the participation in school activities index is of about 0.3 SD, representing a substantial effect size in light of the impact evaluation literature. To better understand this effect, Table 1.2 decomposes the treatment effects on each activity included in the summary index. We find an 11-percentage points (hereafter, pp) increase in participation in parents' cafés on a basis of 26% in the control group (so a small baseline level of participation), but also in events parents attend more frequently such as report card meetings (+6 pp), school

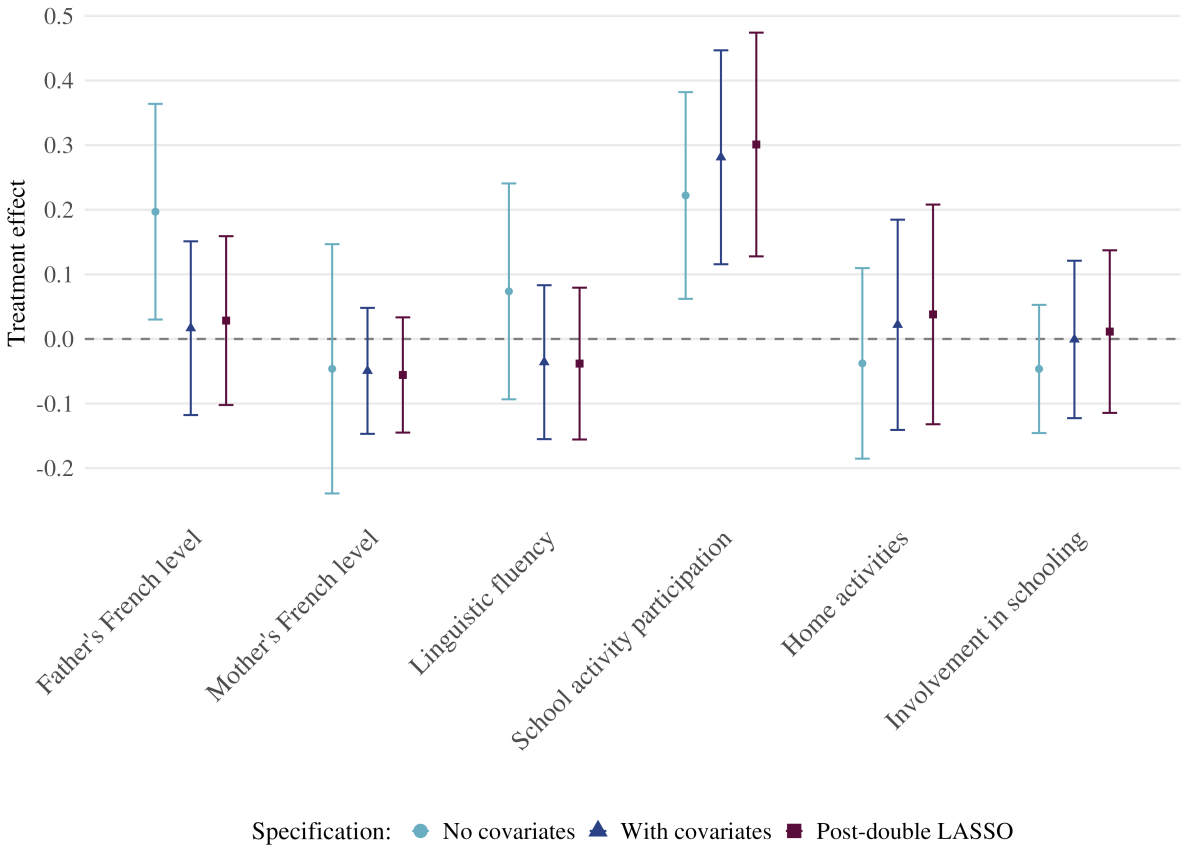

Figure 1.4: ITT Effects on parental language proficiency and engagement in schooling

Notes: Each point is the estimated ITT coefficient for the given outcome and specification. With covariates includes the individual and school-level controls retained by the iterative VIF selection; Post-double LASSO selects from the full pooled set of individual and school-level candidate covariates. All outcomes are standardised on the control-group distribution. Error bars show 90% confidence intervals. Standard errors are clustered at the school level. Full regression tables in Table C1. Sources: parents' survey and DEPP data [ARCHIPEL, 2019-2025].

trips (+7 pp), back-to-school meetings (+8 pp). Individually, only the increase in parents' cafés is statistically significant, but several of the other effects are close to be. Yet, this increased participation in school activities is not accompanied by a significantly greater involvement in their children's schooling (*e.g.*, helping with homework or discussing school matters), nor in educational activities at or outside the home (*e.g.*, reading books, going to the library, or to the cinema). This likely reflects a deficit in economic, socio-cultural, and linguistic capital that the programme does not sufficiently address.

Overall, the specificity of the OEPRE programme seems to lie primarily in the fact that the programme is held at school, thereby bringing parents closer to the school environment, enabling them to better understand how it operates, and fostering increased involvement in the activities it proposes.

Parental Political, Social, and Economic Outcomes. The effects on more indirect outcomes related to the political, social and economic life of parents are presented in Figure 1.5, with more detailed figures also shown in Table C2. The programme shows limited effects on

Table 1.2: ITT estimates on parental participation in school events

School event	Mean (C)	No covariates	With covariates	Post-double LASSO
Parents' coffee morning	0.26	0.06	0.11***	0.12***
		(0.04)	(0.04)	(0.04)
Children's show	0.63	0.07**	0.04	0.06
		(0.04)	(0.05)	(0.04)
Report card meeting	0.71	0.06	0.06	0.05
		(0.04)	(0.05)	(0.05)
School trip	0.46	0.04	0.07	0.07
		(0.05)	(0.05)	(0.05)
Back-to-school meeting	0.67	0.06	0.08	0.08*
		(0.04)	(0.05)	(0.05)
Covariates		No	Yes[†]	BCH-DL[‡]
Obs. range		712-713	656-657	656-657

Notes: This table reports intention-to-treat (ITT) effects of OEPRE school availability on parental participation in five school events. Each outcome is a binary indicator equal to 1 if the respondent reports attending the event, 0 otherwise. Mean (C) is the average participation rate among parents in control schools. Column (1) uses no additional controls; column (2)^{dagger} includes the controls retained by iterative VIF selection (threshold 5 within each block, 10 on the combined set): Father born/raised in FR, Father year of arrival, Father age group, Father region of origin, Mother born/raised in FR, Mother year of arrival, Mother age group, Mother region of origin, Education level, Number of children, Family structure, Residency status, French nationality, Language at arrival (std.), Respondent is mother; school level: UPE2A, ULIS, total pupils, share of girls, teacher age, teacher seniority, teaching staff (FTE), share tenured, share of female teachers, QPV, educational city, REP, REP+, rural, intermediate urban, dense urban, log municipal population, median income, PCS advantaged, PCS intermediate, fragile in French, fragile in maths, satisfactory in maths; column (3)^{ddagger} uses covariates selected by double LASSO (BCH procedure, union of the outcome and treatment equations; selection engine: glmnet::cv.glmnet (validation croisee)) from the full pooled set of individual and school-level candidate covariates, with no prior VIF screening. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Standard errors are clustered at the school level. Sources: parents' survey and DEPP data [ARCHIPEL, 2019-2025].

these outcomes, except for parents' self-esteem: OEPRE led to a statistically significant 0.20 SD increase in self-esteem among parents in the treatment schools.

This sizeable effect reflects increased feelings of pride, greater self-satisfaction, and a heightened awareness of one's own value and capabilities among the parents in treatment schools. This important finding is consistent with the results of the exploratory qualitative study conducted prior to the programme evaluation, which showed that parents most often perceive OEPRE workshops as spaces for overcoming a form of "linguistic shame" fostering self-confidence, and, more broadly, promoting empowerment.

However, OEPRE did not change the other outcomes. We find no effect on civic engagement, trust in institutions, sense of belonging to French society, gender attitudes, social network, and living conditions. These dimensions are less directly targeted by the programme, and it may take more time to change them. It is possible that a longer term time horizon is needed to capture these integration effects.

Children's Academic Learning. The intention-to-treat effects on children's test scores are presented in Figure 1.6, with more detailed figures also shown in Table C3. We find no significant difference in average test scores in French and in Mathematics between children in treatment and control schools. However, the average test scores in French are about 0.10 SD higher in treatment

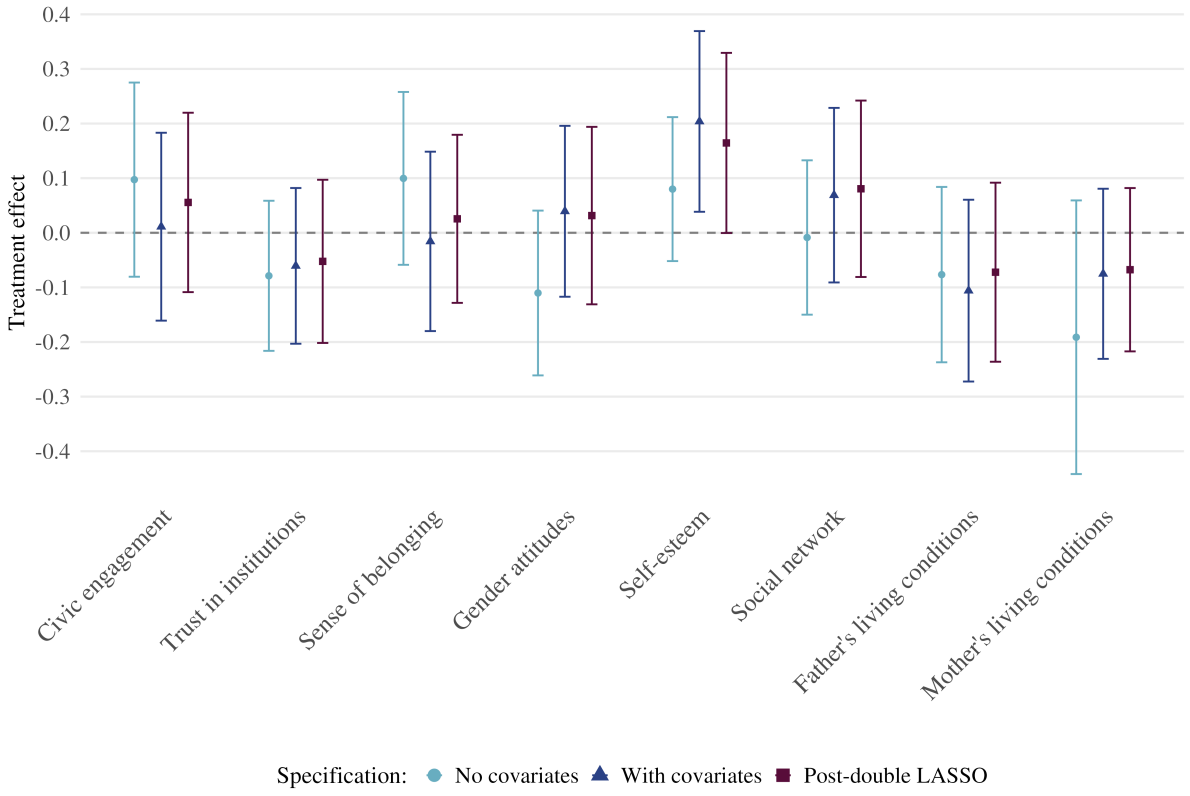

Figure 1.5: ITT Effects on parental political, and economic outcomes

Notes: Each point is the estimated ITT coefficient for the given outcome and specification. With covariates includes the individual and school-level controls retained by the iterative VIF selection; Post-double LASSO selects from the full pooled set of individual and school-level candidate covariates. All outcomes are standardised on the control-group distribution. Error bars show 90% confidence intervals. Standard errors are clustered at the school level. Full regression tables in Table C2. Sources: parents' survey and DEPP data [ARCHIPEL, 2019-2025].

schools compared to control schools, which is not negligible. In fact, educational interventions in high-income countries often fail or have fairly small effects (Kraft, 2020). Unfortunately, the precision of the estimates is not sufficient, so the estimation is not conclusive. The effect of the OEPRE programme on children's test scores in French remains in a grey area.

6.2 Effects on Programme Participants

Figures 1.7 and 1.8 present estimates of OEPRE participation on parental outcomes, under the three identification strategies described in Section 5: the IV estimator which identifies a LATE on compliers, and the CBPS and entropy-balencing reweighting estimators which identify the ATT on actual participants. Detailed estimates and statistics are reported in Tables C4 and C5.

As expected, the IV estimates are about twice as large as the corresponding ITT ones: among compliers (parents who participate in OEPRE only when the programme takes place in their children's school), the programme increases their school activity participation index by 0.57 SD and their self-esteem index by 0.42 SD. The CBPS and Entropy Balancing weighting ATT estimates confirm these results, with significant treatment effects on the same two parent

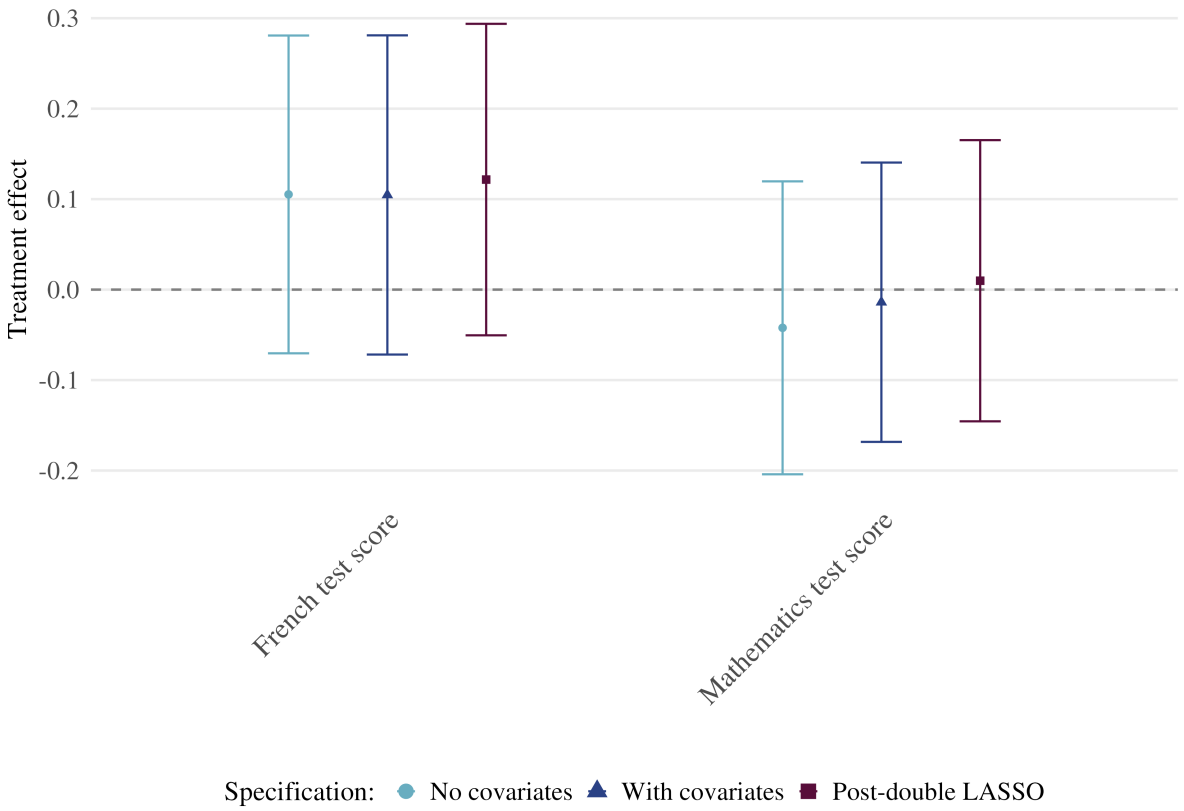

Figure 1.6: ITT effects on children's test scores

Notes: Each point is the estimated ITT coefficient for the given outcome and specification. Outcomes are the child's most recent standardised score in French and in Mathematics (first children, z-scores on the control-school distribution of the same grade); for children of OEPRE participants, the score is strictly posterior to the household's OEPRE starting year. With covariates includes the individual and school-level controls retained by the iterative VIF selection; Post-double LASSO selects from the full pooled candidate set. Error bars show 90% confidence intervals. Standard errors are clustered at the school level. Full regression tables in Table C3. Sources: parents' survey and DEPP data [ARCHIPEL, EVA6, and REPÈRES, 2019-2025].

outcomes: school activity participation (+0.43 SD) and self-esteem (+0.20 SD), along with a marginally significant +0.20 SD effect on fathers' French level.

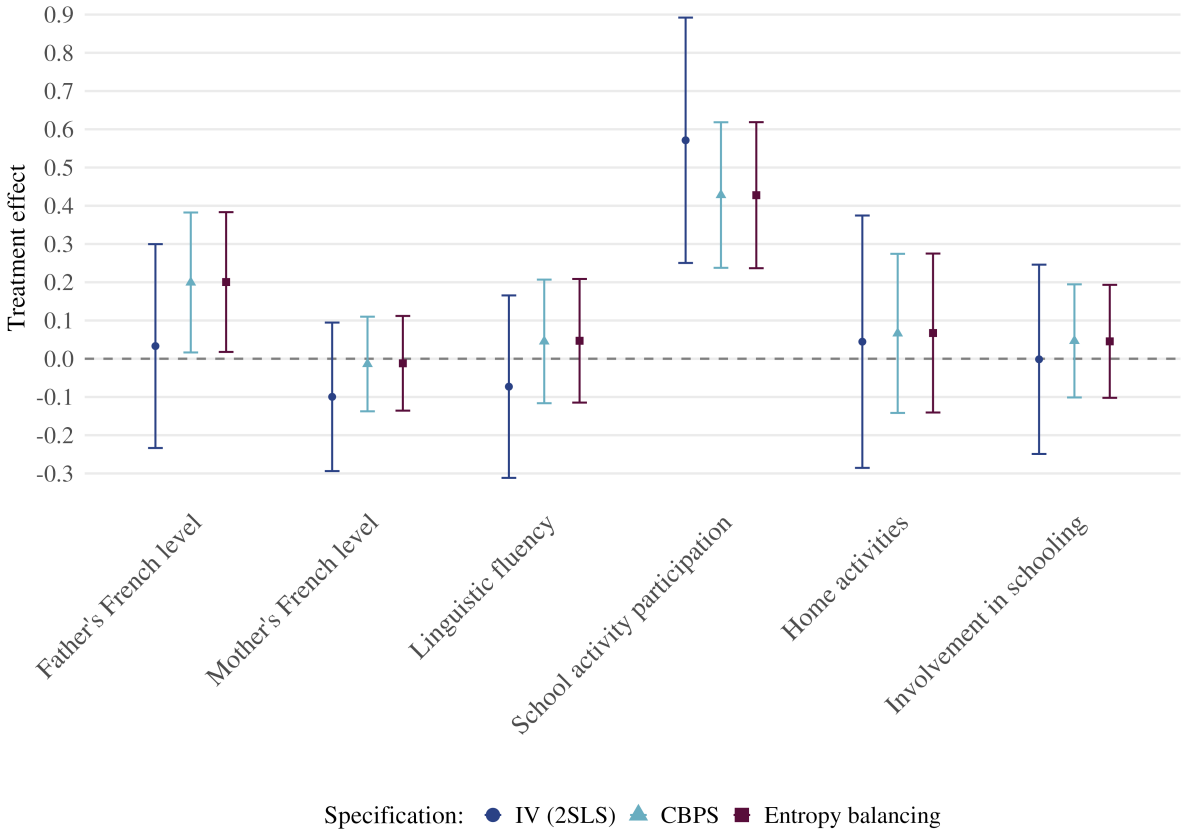

Figure 1.7: Effects of OEPRE Participation on parental language proficiency and engagement in schooling

Notes: Each point shows the estimated effect of OEPRE participation on the given outcome. IV identifies the LATE on compliers using school-level treatment assignment as instrument, with the VIF-selected individual and school-level controls. CBPS and Entropy balancing identify the ATT by reweighting non-participating respondents from control schools to match the household characteristics of participants from treatment schools; the weighted regressions condition on the matching covariates plus the school-level controls. Error bars show 90% confidence intervals. Standard errors are clustered at the school level. Full regression tables in Table C4. Sources: parents' survey and DEPP data [ARCHIPEL, 2019-2025].

ATT points estimates appear lower than the LATE estimates, which may be due to within-school positive externalities: if parents who participate in OEPRE have positive effects on eligible parents in the same school not participating to the programme, then the LATE is overestimated. In that case, the ATT is a more valid measure of the programme's effects than the LATE. In all cases, OEPRE has a large significant impact on participants' involvement in school activities and self-esteem. These are large positive effects that are interesting *per se* since they are likely to improve parents' wellbeing. These effects may also have positive consequences on the children as the economic literature shows that parental involvement in school has positive effects on children's behaviour at school (Avvisati et al., 2010; Avvisati et al., 2014).

In fact, we find that participation in the OEPRE programme actually improve children's literacy. Figure 1.9 and Table C6 present the effects of OEPRE on children's test scores in

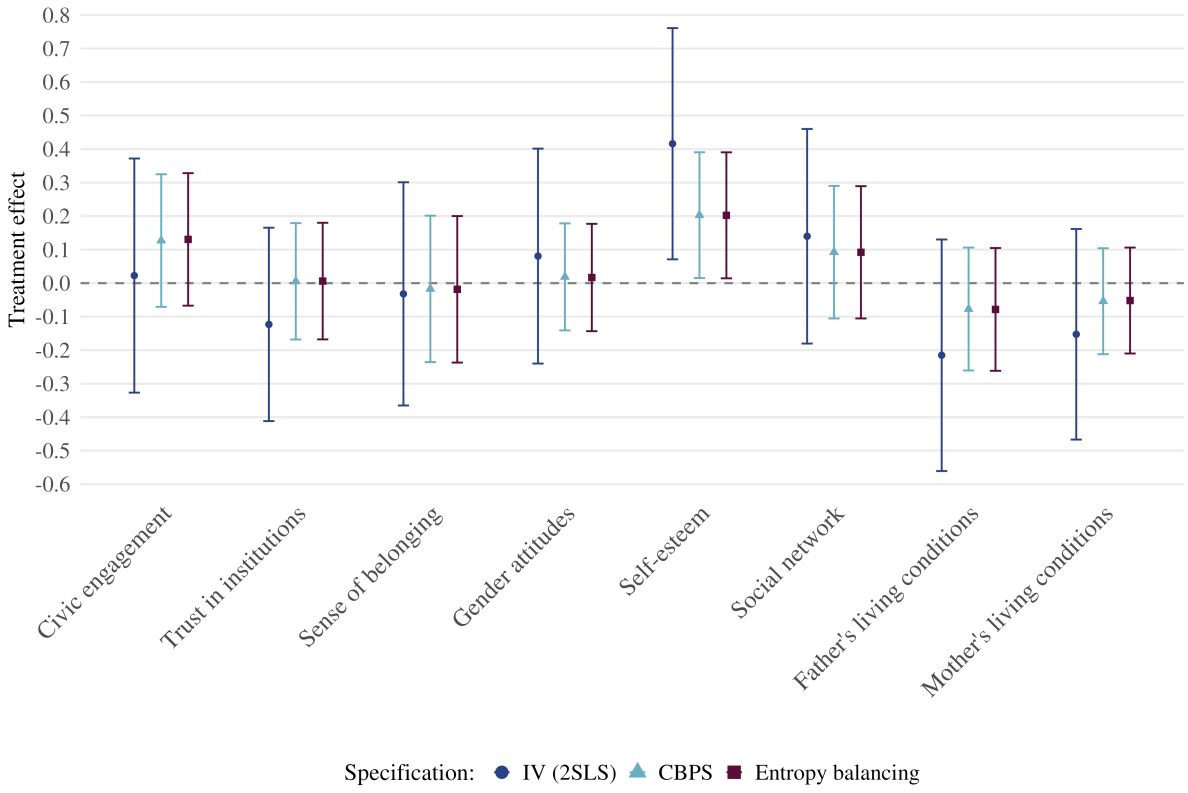

Figure 1.8: Effects of OEPRE participation on parental political, social, and economic outcomes

Notes: Each point shows the estimated effect of OEPRE participation on the given outcome. IV identifies the LATE on compliers using school-level treatment assignment as instrument, with the VIF-selected individual and school-level controls. CBPS and Entropy balancing identify the ATT by reweighting non-participating respondents from control schools to match the household characteristics of participants from treatment schools; the weighted regressions condition on the matching covariates plus the school-level controls. Error bars show 90% confidence intervals. Standard errors are clustered at the school level. Full regression tables in Table C5. Sources: parents' survey and DEPP data [ARCHIPEL, 2019-2025].

French and Mathematics, for those whose parents participate in the programme. The point estimates for French test scores range from +0.25 SD for the IV to +0.30 SD for the weighted estimators, with significance at the 10 percent level for the two ATT estimates. Besides, there is no effect on test scores in Mathematics. The effects on the participants are more than twice as large as the ITT estimates, which again is consistent with the fact that about half of eligible parents participate in OEPRE in schools that offer the programme.

The increase in French test scores is substantial, as the literature shows that only a few programmes have impact on academic learning of that size (Kraft, 2020). The point estimates are unchanged when we include year-of-test fixed effects, grade fixed effects, year-of-test times grade fixed effects, and baseline test scores (with a dummy indicating the missing category for most children since they did not participate in national evaluations before 2023) as additional control variables—the precision of the estimation decreases though, because our sample size is small. Even if the precision of the estimation is not great, the stability of the point estimates provides converging evidence in favour of a positive and sizeable treatment effect on literacy skills for children of participating parents. Therefore, the integration of non-French-speaking

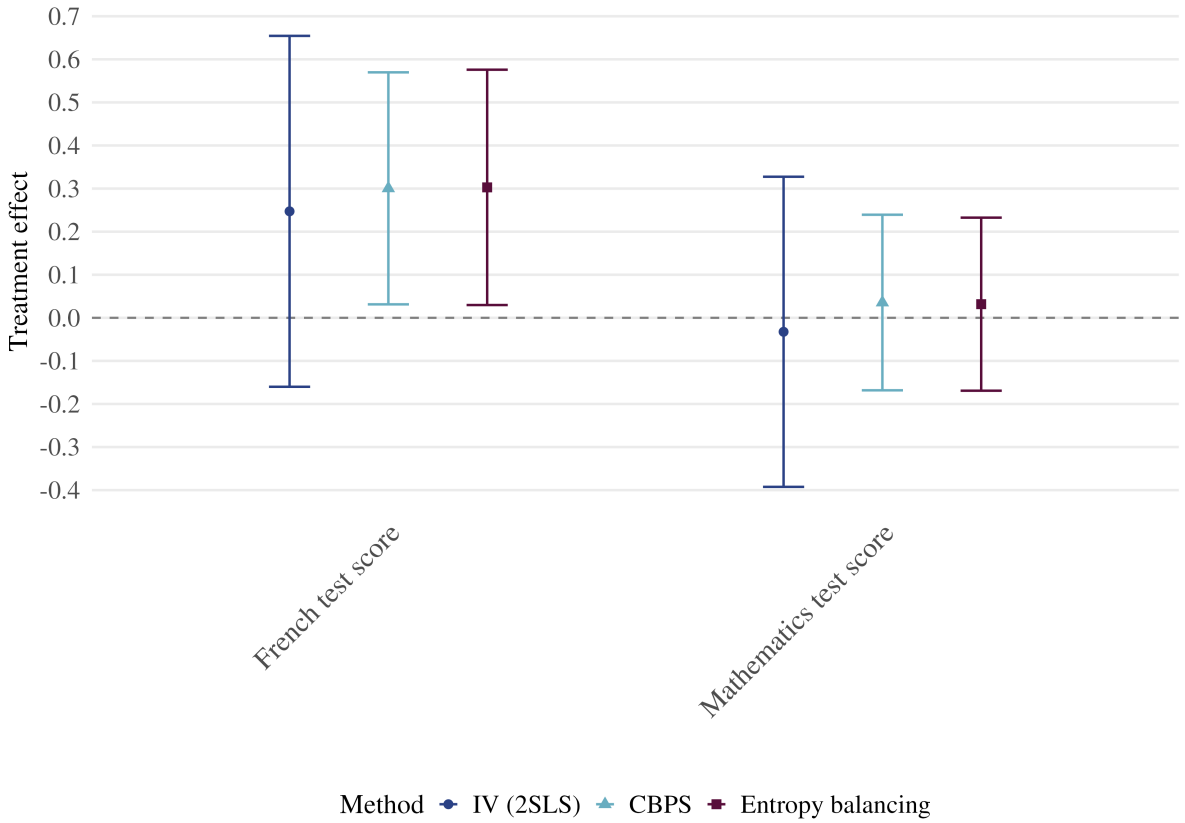

Figure 1.9: Effects of OEPRE participation on children's test scores

Notes: Each point shows the estimated effect of OEPRE participation on the given outcome under three identification strategies. IV identifies the LATE on compliers using school-level treatment assignment as instrument. CBPS and Entropy balancing identify the ATT by reweighting non-participants to match participant characteristics. The treated group consists of children of OEPRE participants from treatment schools; the comparison group consists of children of non-participating households from control schools. Outcomes are the child's most recent standardised score (first children). Error bars show 90% confidence intervals. Standard errors are clustered at the school level. Full regression tables in Table C6. Sources: parents' survey and DEPP data [ARCHIPEL, EVA6, and REPÈRES, 2019-2025].

immigrant parents in schools appears as an efficient policy to reduce the achievement gap between immigrant and native children.

7 Conclusion

This paper evaluates the impact of the Ouvrir l'École aux Parents pour la Réussite des Enfants (OEPRE) programme, a school-based intervention jointly implemented by the French Ministry of Education and the Ministry of the Interior to foster the involvement and integration of non-French-speaking immigrant parents. Using a quasi-experimental design that exploits variation in programme availability across otherwise comparable schools and years, and combining rich administrative data with an original multilingual parent survey, we provide causal evidence on the effects of OEPRE on parental outcomes. The main finding is that OEPRE significantly increases immigrant parents' participation in school activities by about 0.25 standard deviations, a sizeable effect by the standards of the programme evaluation literature. Instrumental variable estimates indicate that, for parents who actually participate as a result of the programme being offered in their children's school, participation in school activities increases by more than half a standard deviation, suggesting that OEPRE succeeds in reducing barriers between immigrant families and schools. Beyond school engagement, the programme has a robust and economically meaningful effect on parents' self-esteem. We estimate an increase of 0.20 standard deviations at the intention-to-treat level and effects exceeding 0.40 standard deviations for participating parents. By creating a supportive collective environment within schools, the programme appears to operate through psychosocial channels—confidence, recognition, and empowerment—that standard language or integration programmes do not explicitly target.

Crucially, the increased participation of parents in school activities and self-esteem translate into 0.30 standard deviations improvement in test scores in French for children of participating parents. This result provides first evidence that the integration of non-French-speaking immigrant parents in schools is an efficient policy to reduce the achievement gap between immigrant and native children. In contrast, we find no detectable impact of the programme on parents' French language proficiency. This null result likely reflects the institutional context: most newly arrived immigrants already benefit from extensive language training through the Republican Integration Contract (CIR), and the additional number of hours offered by OEPRE is modest in comparison. This finding suggests that the distinctive contribution of OEPRE lies less in language acquisition *per se* than in its school-based and community-oriented dimension. Also, we do not find significant short-run effects on other dimensions of integration, such as civic engagement, trust in institutions, sense of belonging, gender attitudes, social networks, or living conditions, outcomes that are plausibly more distant in the programme's theory of change.

Several key takeaways emerge from these results. First, bringing immigrant parents physically and symbolically into schools matters. The school setting not only lowers participation costs but also confers legitimacy on parents' involvement, thereby strengthening school–family links in highly disadvantaged contexts. Second, policies aimed at immigrant integration should explicitly recognise the importance of psychosocial outcomes. Gains in self-esteem and confi-

dence are valuable in their own right and may constitute necessary preconditions for longer-term effects such as labour market participation and children's educational trajectories. Third, the effects on parents translate into increases in children's test scores in French, which suggests that involving parents in schools can foster children's motivation and integration in ways that improve their learning. Finally, the predominance of women among participants underscores the potential of school-based programmes to reach immigrant mothers, a group that remains under-represented in standard adult integration policies despite strong evidence of their central role in child development.

This study nonetheless has limitations as the analysis focuses on relatively short- to medium-run outcomes for parents; some dimensions of integration may require longer horizons to materialise. Even so, this paper contributes to the economics of education and migration by showing that programmes targeting parents rather than children alone can meaningfully shape school engagement and parental well-being in disadvantaged contexts, with downstream effects on academic learning of children in French. In a setting where immigrant integration is both a social and political priority, our findings highlight the potential of schools as key institutional entry points for inclusive and sustainable integration policies.

References

Aizer, Anna (2004). "Home alone: Supervision after school and child behavior". *Journal of public economics* 88.9-10, pp. 1835–1848 (cit. on p. 20).

Alieva, Aigul, Vincent A Hildebrand, and Philippe Van Kerm (2024). "The progression of achievement gap between immigrant and native-born students from primary to secondary education". *Research in Social Stratification and Mobility* 92, p. 100961 (cit. on p. 19).

Antony-Newman, Max (2019). "Parental involvement of immigrant parents: A meta-synthesis". *Educational Review* 71.3, pp. 362–381 (cit. on p. 19).

Attanasio, Orazio, Sarah Cattan, Emla Fitzsimons, Costas Meghir, and Marta Rubio-Codina (2020). "Estimating the production function for human capital: results from a randomized controlled trial in Colombia". *American Economic Review* 110.1, pp. 48–85 (cit. on p. 21).

Avvisati, Francesco, Bruno Besbas, and Nina Guyon (2010). "Parental Involvement in School: A Literature Review". *Revue d'économie politique* 120.5, pp. 759–778 (cit. on pp. 19, 20, 44).

Avvisati, Francesco, Marc Gurgand, Nina Guyon, and Eric Maurin (2014). "Getting Parents Involved: A Field Experiment in Deprived Schools". *Review of Economic Studies* 81, pp. 57–83 (cit. on pp. 21, 44).

Cunha, Flavio and James J Heckman (2008). "Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation". *Journal of human resources* 43.4, pp. 738–782 (cit. on pp. 20, 21).

Cunha, Flavio, James J Heckman, and Susanne M Schennach (2010). "Estimating the technology of cognitive and noncognitive skill formation". *Econometrica* 78.3, pp. 883–931 (cit. on p. 20).

DEPP-MENJ (2020). *67909 élèves allophones nouvellement arrivés en 2018-2019 : neuf sur dix bénéficient d'un soutien linguistique ou d'une scolarité dans un dispositif spécifique, Note d'information de la DEPP, (20.39)*. DEPP-MENJ (cit. on p. 24).

Foged, Mette, Linea Hasager, and Giovanni Peri (2024a). "Comparing the Effects of Policies for the Labor Market Integration of Refugees". *Journal of Labor Economics* 42.S1, S335–S377 (cit. on pp. 19, 21).

Foged, Mette, Linea Hasager, Giovanni Peri, Jacob Nielsen Arendt, and Iben Bolvig (2023). "Intergenerational spillover effects of language training for refugees". *Journal of Public Economics* 220, p. 104840 (cit. on p. 21).

— (2024b). "Language Training and Refugees' Integration". *Review of Economics and Statistics* 106.4, pp. 1157–1166 (cit. on p. 21).

Foged, Mette and Cynthia van der Werf (2023). "Access to language training and the local integration of refugees". *Labour Economics* 84, p. 102366 (cit. on p. 21).

Guryan, Jonathan, Erik Hurst, and Melissa Kearney (2008). "Parental education and parental time with children". *Journal of Economic perspectives* 22.3, pp. 23–46 (cit. on p. 19).

Hainmueller, Jens (2012). "Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies". *Political Analysis* 20.1, pp. 25–46 (cit. on p. 38).

Hasager, Linea (2024). "Does granting refugee status to family-reunified women improve their integration?" *Journal of Public Economics* 234, p. 105119 (cit. on p. 19).

Heller, Blake and Kirsten Slungaard Mumma (2023). "Immigrant Integration in the United States: The Role of Adult English Language Training". *American Economic Journal: Economic Policy* 15.3, pp. 407–437 (cit. on p. 21).

Hill, Nancy E and Lorraine C Taylor (2004). "Parental school involvement and children's academic achievement: Pragmatics and issues". *Current directions in psychological science* 13.4, pp. 161–164 (cit. on p. 19).

Hoover-Dempsey, Kathleen V and Howard M Sandler (1995). "Parental involvement in children's education: Why does it make a difference?" *Teachers college record* 97.2, pp. 310–331 (cit. on p. 19).

Imai, Kosuke and Marc Ratkovic (2014). "Covariate Balancing Propensity Score". *Journal of the Royal Statistical Society Series B: Statistical Methodology* 76.1, pp. 243–263. eprint: https://academic.oup.com/jrsssb/article-pdf/76/1/243/49514325/jrsssb_76_1_243.pdf (cit. on p. 38).

Jung, Eunjoo and Yue Zhang (2016). "Parental involvement, children's aspirations, and achievement in new immigrant families". *The Journal of Educational Research* 109.4, pp. 333–350 (cit. on p. 19).

Kalb, Guyonne and Jan C Van Ours (2014). "Reading to young children: A head-start in life?" *Economics of Education Review* 40, pp. 1–24 (cit. on p. 20).

Kim, Sungwon (2022). "Fifty years of parental involvement and achievement research: A second-order meta-analysis". *Educational Research Review* 37, p. 100463 (cit. on p. 19).

Kraft, Matthew A (2020). "Interpreting effect sizes of education interventions". *Educational Researcher* 49.4, pp. 241–253 (cit. on pp. 20, 42, 45).

Lareau, Annette (1987). "Social class differences in family-school relationships: The importance of cultural capital". *Sociology of education*, pp. 73–85 (cit. on p. 19).

Lavy, Victor, Giulia Lotti, and Zizhong Yan (2022). "Empowering mothers and enhancing early childhood investment: Effect on adults' outcomes and children's cognitive and noncognitive skills". *Journal of Human Resources* 57.3, pp. 821–867 (cit. on p. 19).

Lochmann, Alexia, Hillel Rapoport, and Biagio Speciale (2019). "The effect of language training on immigrants' economic integration: Empirical evidence from France". *European Economic Review* 113, pp. 265–296 (cit. on p. 21).

Ministère de l'Éducation nationale (2025). *Bulletin officiel n°38 du 9 octobre 2025 - Dispositif Ouvrir l'École aux parents pour la réussite des enfants (OEPRE)*. https://www.education.gouv.fr/bo/2025/Hebdo38/MENE2525824C (cit. on p. 20).

Nations, United (2024). *International Migrant Stock 2024: Key facts and figures*. Report 13. UN DESA/POP/2024/DC (cit. on p. 19).

OECD (2025). *Résultats du PISA 2022 (Volume I) : Apprentissage et équité dans l'éducation, PISA*. OECD (cit. on p. 19).

Rosenberg, Morris (1965). "Rosenberg self-esteem scale". *Journal of Religion and Health* (cit. on p. 61).

Sarvimäki, Matti and Kari Hämäläinen (2016). "Integrating immigrants: The impact of restructuring active labor market programs". *Journal of Labor Economics* 34.2, pp. 479–508 (cit. on p. 21).

Todd, Petra E and Kenneth I Wolpin (2007). "The production of cognitive achievement in children: Home, school, and racial test score gaps". *Journal of Human capital* 1.1, pp. 91–136 (cit. on p. 20).

Tucci, Ingrid, Jean-Luc Primon, and Mathieu Ichou (2026). "Diplômes des immigrés et des descendants d'immigrés: progrès intergénérationnels et persistance des inégalités". *Diversité des populations et inégalités en France. L'enquête Trajectoires et Origines* (cit. on p. 19).

Turney, Kristin and Grace Kao (2009). "Barriers to school involvement: Are immigrant parents disadvantaged?" *The Journal of Educational Research* 102.4, pp. 257–271 (cit. on p. 19).

Welsch, David M and David M Zimmer (2008). "After-school supervision and children's cognitive achievement". *The BE Journal of Economic Analysis & Policy* 8.1 (cit. on p. 21).

Appendix

A Additional Tables and Figures

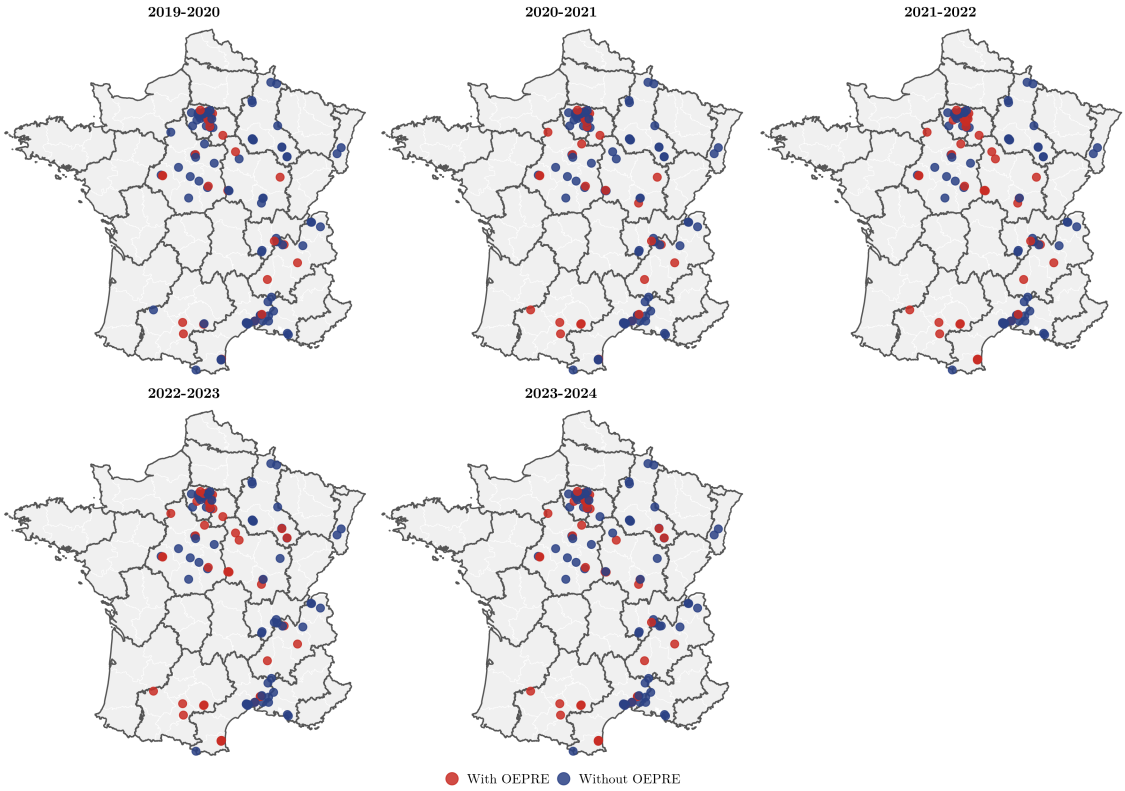

Figure A1: Geographic Distribution of OEPRE Schools, 2019–2024

Notes: Each panel shows schools hosting an OEPRE workshop (red circles, *With OEPRE*) and schools without (blue circles, *Without OEPRE*) for each academic year from 2019–2020 to 2023–2024. The sample comprises 148 schools whose OEPRE status evolves over time, from 42 schools with OEPRE in 2019–2020 to 68 in 2022–2023. The 2022–2023 panel corresponds to Figure 1.2 in the main text.

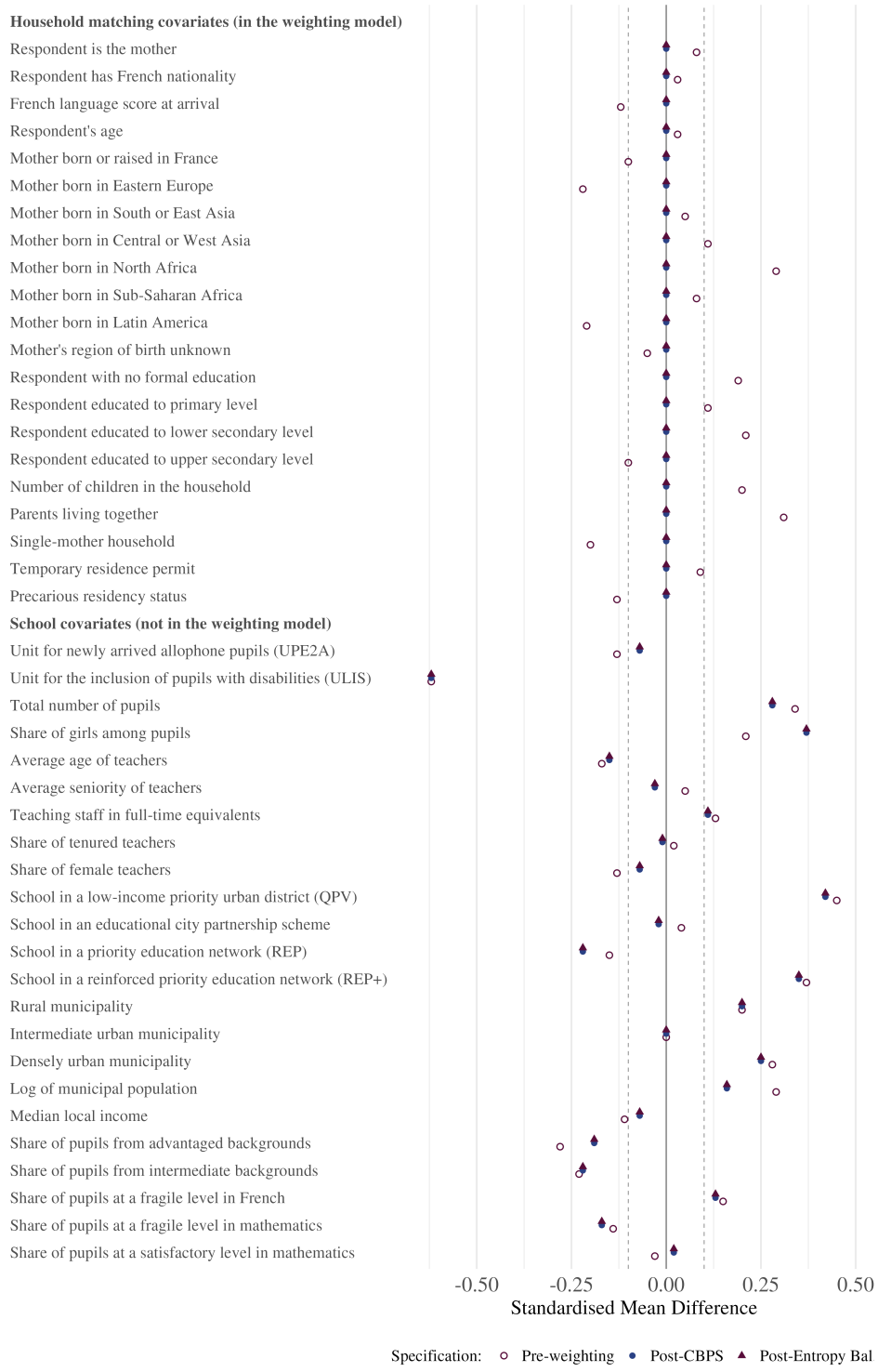

Figure A2: Balancing between treated and non-treated parents for parental outcomes, before and after CBPS and entropy balancing score-based reweighting

Notes: Each point shows the standardised mean difference for a given covariate between OEPRE participants from treatment schools and non-participating respondents from control schools, before and after reweighting, in the sample used for the weighted ATT estimates. The weights balance the household matching covariates of the upper panel; the school covariates of the lower panel are not part of the weighting model and are shown to assess how school environments compare, before and after the household weights are applied; they are controlled for in the doubly robust regressions. Dashed lines indicate the ±0.1 acceptability threshold. Sources: parents' survey and DEPP data [ARCHIPEL, 2019-2025].

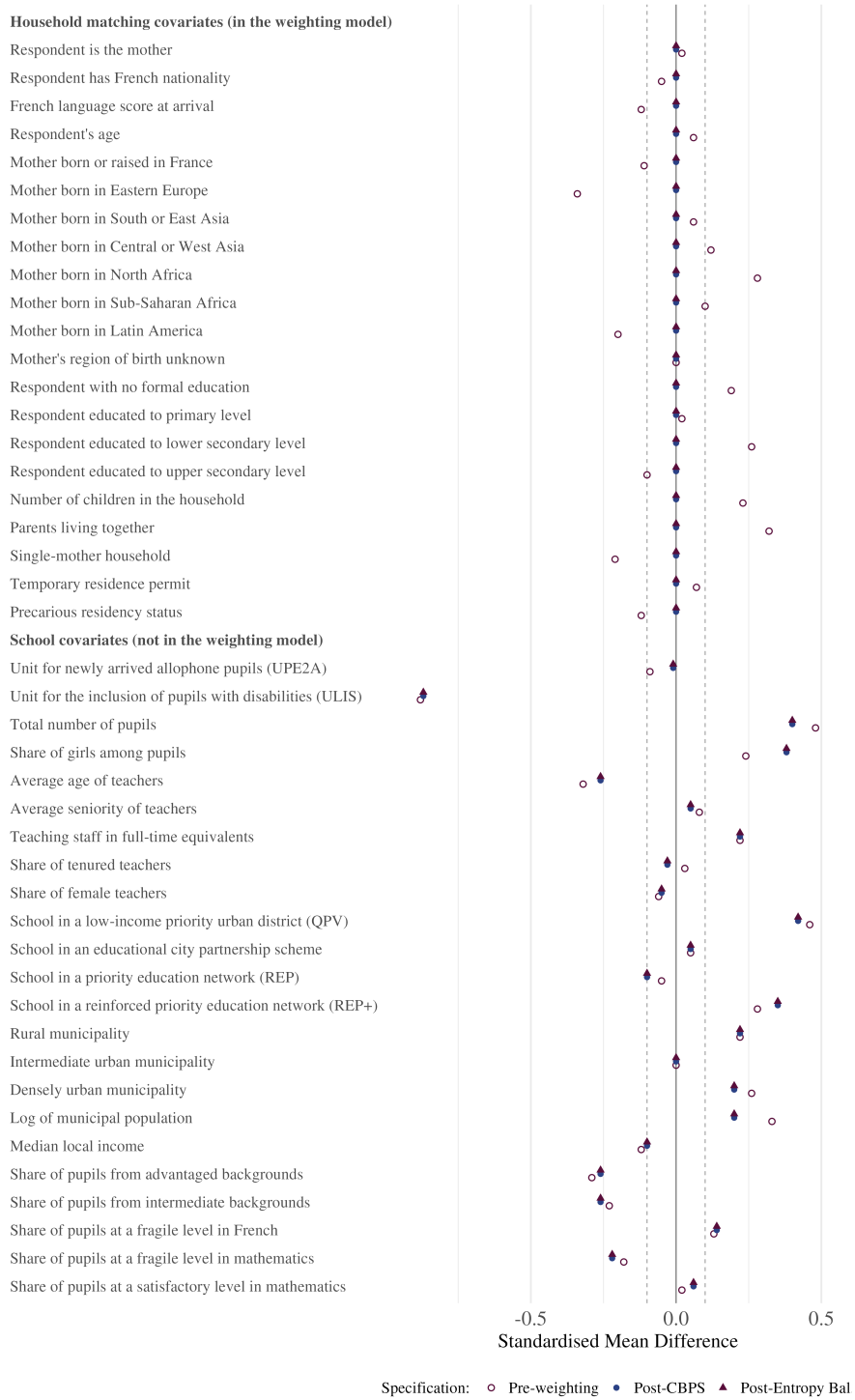

Figure A3: Balancing between treated and non-treated parents for children's outcomes, before and after CBPS and entropy balancing score-based reweighting

Notes: Each point shows the standardised mean difference for a given covariate between children of OEPRE participants from treatment schools and children of non-participating households from control schools, before and after reweighting, in the sample used for the ATT estimates on children's test scores. The weights balance the household matching covariates of the upper panel; the school covariates of the lower panel are not part of the weighting model and are shown to assess how school environments compare, before and after the household weights are applied; they are controlled for in the doubly robust regressions of the specification with school covariates. Dashed lines indicate the ±0.1 acceptability threshold. Sources: parents' survey and DEPP data [ARCHIPEL, 2019-2025].

Table A1: Baseline Characteristics of Sampled Parents

	Full sample		OEPRE participants		Difference		
Variable	N	Mean	N	Mean	Diff.	SE	p-value
Panel A: Respondent							
Household characteristics							
Age	700	39.400	266	39.466	0.066	0.513	0.897
Mothers among respondents	713	0.745	270	0.763	0.018	0.031	0.552
Number of children	711	2.734	269	2.766	0.032	0.083	0.704
Youngest child born outside France	713	0.675	270	0.615	-0.060	0.034	0.083
Parents living together	712	0.795	269	0.859	0.064	0.026	0.015
Single mother	712	0.167	269	0.126	-0.041	0.025	0.099
Other arrangement	712	0.038	269	0.015	-0.023	0.010	0.025
Presence in France							
French nationality	710	0.049	269	0.041	-0.008	0.015	0.564
French proficiency at arrival	687	1.622	262	1.533	-0.089	0.057	0.121
Residence status							
Stable residency	688	0.302	258	0.298	-0.004	0.033	0.908
Temporary permit	688	0.222	258	0.248	0.026	0.031	0.412
Precarious status	688	0.475	258	0.453	-0.022	0.036	0.550
Panel B: Father							
Presence in France							
Age	603	42.828	231	42.974	0.146	0.584	0.802
Born/raised in France	685	0.067	261	0.088	0.021	0.020	0.295
Year of arrival	565	2014.588	223	2014.354	-0.233	0.808	0.773
Origins							
From Western Europe/USA	632	0.103	247	0.081	-0.022	0.021	0.302
From Eastern Europe	632	0.112	247	0.073	-0.039	0.021	0.058
From South/East Asia	632	0.179	247	0.198	0.020	0.030	0.509
From Central/Western Asia	632	0.161	247	0.190	0.029	0.029	0.320
From North Africa	632	0.291	247	0.336	0.045	0.035	0.202
From Sub-Saharan Africa	632	0.087	247	0.085	-0.002	0.021	0.924
From Latin America	632	0.066	247	0.036	-0.030	0.016	0.054
Education							
No education	169	0.065	59	0.085	0.020	0.041	0.635
Primary education	169	0.053	59	0.051	-0.002	0.034	0.943
Lower secondary	169	0.154	59	0.186	0.033	0.058	0.577
Upper secondary	169	0.361	59	0.305	-0.056	0.071	0.433
Higher education	169	0.367	59	0.373	0.006	0.074	0.935
Panel C: Mother							
Presence in France							
Age	693	38.023	265	38.106	0.083	0.449	0.854
Born/raised in France	711	0.034	270	0.030	-0.004	0.012	0.739
Year of arrival	697	2017.779	267	2018.004	0.225	0.413	0.586
Origins							
From Western Europe/USA	707	0.071	269	0.033	-0.037	0.015	0.011
From Eastern Europe	707	0.133	269	0.089	-0.044	0.022	0.043
From South/East Asia	707	0.168	269	0.197	0.029	0.028	0.307
From Central/Western Asia	707	0.156	269	0.182	0.027	0.027	0.330
From North Africa	707	0.286	269	0.349	0.064	0.034	0.059
From Sub-Saharan Africa	707	0.096	269	0.097	0.000	0.021	0.982
From Latin America	707	0.091	269	0.052	-0.038	0.017	0.027
Education							
No education	511	0.074	192	0.104	0.030	0.025	0.234
Primary education	511	0.068	192	0.083	0.015	0.023	0.518
Lower secondary	511	0.168	192	0.224	0.056	0.034	0.107
Upper secondary	511	0.313	192	0.286	-0.027	0.039	0.491
Higher education	511	0.376	192	0.302	-0.074	0.040	0.063

Notes: This table reports baseline characteristics of all sampled parents (full sample, N = 713) and OEPRE participants (OEPRE participants, N = 270). All variables are binary (reported as proportions) except: age (years), year of arrival (year), number of children (count), and French proficiency at arrival (average of four self-assessed items – reading, writing, speaking, understanding – on a 1–4 scale where 1 = not at all and 4 = very well). Origin dummies exclude cases where the parent's birthplace was not reported.Diff. reports the difference in means between OEPRE participants and the full sample. Standard errors from a two-sample t-test with unequal variances.

Table A2: Balance Check: Treatment vs Control School Characteristics

	Control		Treatment		Difference		
Variable	N	Mean	N	Mean	Diff.	SE	p-value
School structure							
UPE2A (language support class)	53	0.491	95	0.484	-0.006	0.086	0.941
ULIS (disability support class)	53	0.434	95	0.263	-0.171	0.082	0.040
Number of classes	53	10.811	94	11.649	0.838	0.670	0.213
Total pupils	53	202.491	94	205.128	2.637	12.124	0.828
Share of girls	53	0.480	94	0.487	0.006	0.006	0.309
Teaching staff characteristics							
Mean teacher age	53	40.766	94	40.326	-0.440	0.740	0.553
Mean teacher seniority	53	6.000	94	5.855	-0.145	0.380	0.703
Teaching staff (FTE)	53	12.952	94	13.530	0.578	0.731	0.430
Share of tenured teachers	53	0.798	94	0.802	0.004	0.021	0.833
Share of female teachers	53	0.852	94	0.831	-0.021	0.019	0.275
Urban policies							
QPV (priority urban area)	53	0.528	95	0.737	0.209	0.083	0.013
Educational city	53	0.415	95	0.411	-0.005	0.085	0.957
Priority education status							
Outside priority education	53	0.415	95	0.232	-0.184	0.081	0.025
REP (priority education)	53	0.396	95	0.463	0.067	0.085	0.432
REP+	53	0.189	95	0.305	0.117	0.072	0.108
Density characteristics							
Rural	53	0.000	95	0.021	0.021	0.015	0.158
Intermediate urban	53	0.283	95	0.284	0.001	0.078	0.988
Dense urban	53	0.717	95	0.695	-0.022	0.078	0.777
School economic environment							
Municipal population (2020)	53	99200.000	95	287735.968	188535.968	69301.416	0.007
Median income (municipality)	53	17811.698	95	16042.211	-1769.488	1232.652	0.153
Mean School's Social Position Score (IPS)	53	87.961	94	82.236	-5.725	2.643	0.032
Pupils' socioeconomic background							
Share highly advantaged pupils	53	0.132	94	0.089	-0.043	0.023	0.068
Share advantaged pupils	53	0.092	94	0.072	-0.021	0.007	0.007
Share intermediate pupils	53	0.236	94	0.251	0.015	0.014	0.274
Share disadvantaged pupils	53	0.540	94	0.588	0.048	0.031	0.118
Standardised assessments, French (pooled across grades)							
'At risk', French (pooled)	53	0.160	95	0.166	0.005	0.010	0.584
'Fragile', French (pooled)	53	0.221	95	0.231	0.009	0.007	0.179
'Satisfactory', French (pooled)	53	0.618	95	0.603	-0.015	0.015	0.332
Standardised assessments, maths (pooled across grades)							
'At risk', maths (pooled)	53	0.183	95	0.187	0.004	0.009	0.668
'Fragile', maths (pooled)	53	0.203	95	0.204	0.001	0.005	0.848
'Satisfactory', maths (pooled)	53	0.614	95	0.609	-0.005	0.012	0.697

Notes: The difference between treatment and control school means is estimated by OLS regression with standard errors clustered at the school (UAI) level. Assessment scores are pooled across 1st, 2nd, and 4th grades (average of grade-level shares). Scope: treatment and control schools in the OEPRE evaluation sample, years 2022–2023. Sources: ARCHIPEL, REPÈRES, INSEE, authors' calculations.

Table A3: Balance Checks: Parent Characteristics

		Means		Difference		
Variable	N	Control (N=265)	Treatment (N=448)	Diff.	SE	p-value
Panel A: Respondent						
Household characteristics						
Age	700	39.31	39.452	0.142	0.697	0.838
Mothers among respondents	713	0.736	0.75	0.014	0.037	0.703
Number of children	711	2.619	2.803	0.184	0.109	0.092
Youngest child born outside France	713	0.762	0.623	-0.139	0.058	0.017
Parents living together	712	0.77	0.81	0.04	0.035	0.257
Single mother	712	0.185	0.157	-0.028	0.038	0.456
Other arrangement	712	0.045	0.034	-0.012	0.017	0.491
Presence in France						
French nationality	710	0.038	0.056	0.018	0.019	0.344
French proficiency at arrival	687	1.662	1.599	-0.063	0.094	0.502
Residence status						
Stable residency	688	0.27	0.322	0.051	0.043	0.236
Temporary permit	688	0.216	0.226	0.01	0.038	0.796
Precarious status	688	0.514	0.452	-0.061	0.053	0.246
Panel B: Father						
Presence in France						
Age	603	42.449	43.053	0.604	0.81	0.456
Born/raised in France	685	0.047	0.079	0.033	0.017	0.061
Year of arrival	565	2015.735	2013.904	-1.831	1.238	0.14
Origins						
From Western Europe/USA	632	0.114	0.096	-0.018	0.047	0.693
From Eastern Europe	632	0.136	0.098	-0.037	0.035	0.285
From South/East Asia	632	0.136	0.205	0.069	0.042	0.1
From Central/Western Asia	632	0.157	0.164	0.007	0.049	0.882
From North Africa	632	0.297	0.288	-0.009	0.056	0.876
From Sub-Saharan Africa	632	0.085	0.088	0.004	0.028	0.896
From Latin America	632	0.076	0.061	-0.016	0.024	0.512
Education						
No education	169	0.06	0.069	0.009	0.038	0.814
Primary education	169	0.075	0.039	-0.035	0.044	0.417
Lower secondary	169	0.149	0.157	0.008	0.055	0.891
Upper secondary	169	0.313	0.392	0.079	0.077	0.31
Higher education	169	0.403	0.343	-0.06	0.08	0.453
Panel C: Mother						
Presence in France						
Age	693	37.839	38.13	0.291	0.686	0.671
Born/raised in France	711	0.03	0.036	0.005	0.014	0.698
Year of arrival	697	2018.696	2017.233	-1.463	0.982	0.137
Origins						
From Western Europe/USA	707	0.111	0.047	-0.063	0.04	0.117
From Eastern Europe	707	0.16	0.117	-0.043	0.036	0.225
From South/East Asia	707	0.13	0.191	0.061	0.039	0.116
From Central/Western Asia	707	0.153	0.157	0.005	0.049	0.924
From North Africa	707	0.267	0.297	0.029	0.054	0.582
From Sub-Saharan Africa	707	0.08	0.106	0.025	0.028	0.36
From Latin America	707	0.099	0.085	-0.014	0.026	0.6
Education						
No education	511	0.062	0.082	0.02	0.025	0.415
Primary education	511	0.046	0.082	0.036	0.023	0.114
Lower secondary	511	0.144	0.183	0.039	0.032	0.227
Upper secondary	511	0.314	0.312	-0.002	0.046	0.963
Higher education	511	0.433	0.341	-0.092	0.055	0.094

Notes: This table reports pre-determined characteristics of parents in treatment and control schools. All variables are binary (reported as proportions) except: age (years), year of arrival (year), number of children (count), and French proficiency at arrival (average of four self-assessed items—reading, writing, speaking, understanding—on a 1–4 scale where 1 = not at all and 4 = very well). Origin dummies exclude cases where the parent's birthplace was not reported (question not asked when the other parent is absent from the household). Diff. reports the difference in means (treatment minus control). Education dummies are assigned based on the respondent's gender. Standard errors are clustered at the school level.

B Construction of the Variables and Indices

All outcomes indices are constructed as summary indices. For each index, each component item is first standardised using the control group mean and standard deviation. The index is then computed as the unweighted average of the standardised items for each respondent, using all available (non-missing) items. Finally, the resulting average is itself standardised on the control group distribution. Treatment effects are therefore expressed as a percentage of a standard deviation relative to the control group.

B.1 Outcomes

French proficiency Respondents were asked to assess the current French proficiency of each parent across four dimensions by answering *"First, can you tell me how well you know French French for ..."*:

1. *"Reading?"*
2. *"Writing?"*
3. *"Speaking?"*
4. *"Understanding?"*

Each item was rated on a four-point scale (1) "very well", (2) "well", (3) "a little", (4) "not at all". The scale was reversed so that higher values indicate better proficiency. "Do not know" and "not administrable" responses were coded as missing. Separate indices were constructed for fathers and mothers.

Linguistic fluency (respondent) This index captures the respondent's overall ease with the French language, combining eight items from different sections of the survey:

1. *"Do you have difficulty requesting information or an administrative service?"* recoded so that 1 indicates having no difficulty.
2. *"Have you obtained a French language diploma (e.g., DELF, DILF, DALF, TCF)?"* a dummy variable indicating if the respondent has a French language diploma diploma. "Don't know" and "Not administrable" responses were coded as 0.
3. The language in which the interview was conducted was derived into a categorical variable taking 2 if the interview was exclusively in French, 1 if partially in French, and 0 if in another language.
4. The extent of French spoken at home with children is derived from the question *"What language(s) do you speak at home with your children?"* for which options were: "exclusively French," "mostly French," "two languages equally," "mostly another language," and "exclusively another language." Recoded to an ordinal scale from 0 (exclusively another language) to 4 (exclusively French).

5. Self-reported improvement in French for reading, writing, speaking, and understanding are defined as the difference between current proficiency and proficiency upon arrival in France in the considered dimension on the reversed four-point scale described above. A positive value thus indicates improvement. Set to missing when either the current or arrival level was unavailable.

School event participation Respondents were asked: *"This year, have you or the other parent participated in:"* followed by five school events:

1. *"The back-to-school meeting"*
2. *"At least one school trip"*
3. *"Report card meeting"*
4. *"A children's show"*
5. *"A parents' coffee meeting"*

Each item was binary (yes/no), with an additional option "there was none at our school." The latter was cross-checked against responses from other parents in the same school: when other parents reported attending, the "there was none" response was recoded to "no." All remaining "there was none," "don't know," and "not administrable" responses were coded as 0.

Home activities with children Respondents were asked about the frequency of activities with their children. The introductory text was: *"We would also like to know what your child likes to do at home, and what you or the other parent do together when you have time."* This was followed by: *"Over the past month, have you:"* and seven items:

1. *"Read or looked at books with your child, regardless of the language?"*
2. *"Talked about book stories with your child?"*
3. *"Done cooking, crafts, or other manual activities with your child?"*
4. *"Gone to the swimming pool or the beach with your children?"*
5. *"Gone to the cinema?"*
6. *"Gone to the library?"*
7. *"Gone to the ice rink?"*

Each item was rated on a four-point frequency scale: (1) "never," (2) "once or twice a month," (3) "once or twice a week," (4) "every day or nearly." "Don't know" and "not administrable" responses were coded as 1 ("never").

Involvement in schooling This index combines three items measuring parental engagement in the child's school life.

Respondents were first asked: *"Do you or the other parent ever talk with your child about..."* followed by two items:

1. *"What he/she does in class? If yes, how often?"*
2. *"How things are going at school with other children? If yes, how often?"*

Each item was rated on a four-point frequency scale: (1) "every day or nearly," (2) "once or twice a week," (3) "once or twice a month," (4) "never." The scale was reversed so that higher values indicate more frequent discussion (4 = every day, 1 = never). "Don't know" and "not administrable" responses were coded as 1 ("never").

The third item asked: *"Does your child most often do homework alone or with someone during the week?"* Response options were: (1) "completely alone," (2) "alone but someone is available to help if needed," (3) "together with someone," (4) "does homework at school," (5) "no homework." This was recoded into an ordinal scale of parental involvement: 1 = alone, 2 = homework at school or no homework, 3 = parent helps if needed, 4 = parent accompanies systematically. "Don't know" and "not administrable" responses were handled case-by-case based on the child's grade level and family situation.

Civic engagement Respondents were asked: *"There are different ways to try to improve things or to prevent things from getting worse. In the past 12 months, have you done any of the following?"*

1. *"Signed a petition?"*
2. *"Voted in an election?"*
3. *"Taken part in a lawful public demonstration?"*
4. *"Gone on strike?"*

Each item was binary (yes/no). "Don't know" and "not administrable" responses were coded as 0 (did not participate). These four items are combined into a summary index; higher values indicate greater civic participation.

Trust in institutions Respondents were asked: *"For each of the following institutions, tell me whether you trust them or not, using a scale from 0 to 10."* :

1. The national education system
2. The press
3. The police

4. Businesses
5. The government
6. The justice system
7. Employers
8. The internet
9. Banks

Each item was rated on a 0–10 scale. "Don't know" responses were coded as 5 (midpoint of the scale); "prefer not to answer" was coded as 0; "not administrable" was coded as missing. These nine items are combined into a summary index; higher values indicate greater institutional trust.

Sense of belonging Respondents were told: *"I am going to read you some statements about how you personally feel in France. Please tell me whether you strongly agree, somewhat agree, somewhat disagree, or strongly disagree."* Three items followed:

1. *"I feel at home in France"*
2. *"I am seen as French"*
3. *"I feel French"*

Items were rated on a four-point agreement scale: (1) "strongly agree," (2) "somewhat agree," (3) "somewhat disagree," (4) "strongly disagree." The scale was reversed to a five-point scale (5 = strongly agree, 1 = strongly disagree) so that higher values indicate a stronger sense of belonging. "Don't know" and "prefer not to answer" responses were placed at the midpoint (3); "not administrable" was coded as missing. These three items are combined into a summary index; higher values indicate a stronger sense of belonging.

Gender attitudes This index captures attitudes toward gender equality through three matched pairs of statements. Respondents were told: *"I am going to read you some sentences. Please tell me whether you strongly agree, somewhat agree, somewhat disagree, or strongly disagree."*

Pro-equality statements:

1. *"It is important that girls have access to education"*
2. *"It is important that women be able to work and earn a wage"*
3. *"It is important that women participate in household decision-making"*

Pro-male-priority statements:

4. *"It is more important for boys than girls to have access to education"*

5. *"It is more important for men than women to work and earn a wage"*
6. *"It is more important for men than women to participate in household decision-making"*

All items used a four-point agreement scale: (1) "strongly agree" to (4) "strongly disagree." Pro-equality items (1–3) were reversed so that higher values indicate stronger agreement with equality. Pro-male-priority items (4–6) were left unchanged, since higher values (disagreement with male priority) already indicate more egalitarian attitudes. "Don't know," "prefer not to answer," and "not administrable" responses were coded as missing. All six items are combined into a summary index; higher values indicate more egalitarian gender attitudes.

Self-esteem This index is adapted from the Rosenberg Self-Esteem Scale (Rosenberg, 1965). Respondents were asked to rate their agreement with five statements:

1. *"I think I am a person of worth, at least equal to anyone else"*
2. *"I can do things as well as most people"*
3. *"I see little reason to be proud of myself"* (reverse-worded)
4. *"I have a positive attitude toward myself"*
5. *"Overall, I am satisfied with myself"*

Items used the same four-point agreement scale as above. Positively-worded items (1, 2, 4, 5) were reversed to a five-point scale (5 = strongly agree, 1 = strongly disagree). For the negatively-worded item (3), strong disagreement–indicating high self-esteem–receives the highest value. "Don't know" responses were placed at the midpoint (3); "prefer not to answer" and "not administrable" were coded as missing. These five items are combined into a summary index; higher values indicate higher self-esteem.

Social network This index captures the extent of the respondent's social network through three items:

1. *"In the past 15 days, aside from your family, how many friends have you met with?"* — Response options: "more than 10," "5 to 9," "1 to 4," "none." Reversed so that higher values indicate a larger network (3 = more than 10, 0 = none). "Don't know" and "not administrable" were coded as missing.
2. *"What share of your friends is from the same country of origin as you?"* — Response options: "less than half," "half," "more than half." Recoded to a 0–2 scale (0 = less than half, 2 = more than half). "Don't know" and "not administrable" were coded as missing.
3. *"Are you a member of any association, regardless of its purpose? E.g., neighbourhood association, solidarity association, cultural association."* — Binary: 1 = yes, 0 = no. "Don't know" was coded as missing.

These three items are combined into a summary index; higher values indicate a larger and more active social network.

Living conditions (father/mother) This index combines five indicators of material well-being:

1. **Housing comfort** — Derived from *"What type of housing do you live in?"* Response options included private rental, social housing, homeownership, hotel, shelter, informal housing, and staying with family or friends, among others. Coded as 1 for stable housing (private rental, social housing, or homeownership); 0 for precarious housing (hotel, shelter, informal housing, staying with family or friends).

2. **Employment status** — Derived from *"What is your employment situation this month?"* Response options included salaried employment, interim/seasonal work, training, independent work, business creation, unemployment, retirement, and awaiting rights. Coded as 1 if employed (salaried, interim, training, independent, or other activity); 0 if unemployed, retired, or awaiting rights.

3. **Type of contract** — Derived from employment status and *"You indicated being in salaried employment this month. Is it a fixed-term or permanent contract?"* Ordinal scale: 0 = inactive, 1 = precarious (interim, seasonal), 2 = fixed-term (CDD), 3 = permanent (CDI).

4. **Weekly hours worked** — Sum of declared hours across all activity types. Set to 0 for inactive respondents.

5. **ISEI occupational prestige** — Derived from *"What is your occupation?"* Free-text responses were standardised, coded to the ISCO-08 four-digit classification via exact and fuzzy string matching, and converted to ISEI scores. Inactive respondents were assigned the minimum ISEI value of 17, following OECD conventions.

These five items are combined into a summary index; higher values indicate better material conditions. Since items are measured on heterogeneous scales (binary, ordinal, and continuous), the item-level standardization places them on a common metric before averaging. Separate indices were constructed for fathers and mothers.

B.2 Control variables

The following covariates are included in all specifications with controls. All categorical variables enter the regressions as factor indicators with explicit reference categories.

French proficiency at arrival Respondents who did not grow up in France were asked: *"When you arrived in France, what was your knowledge of French for: reading? writing? speaking? understanding?"* Each item was rated on the same four-point scale as current proficiency (reversed so that higher values indicate better proficiency). The index is constructed using the same summary index method as the outcome variables. This variable is used exclusively as a covariate.

Respondent characteristics

- **Respondent is female** — Binary indicator derived from the respondent's relationship to the child (mother vs. father).
- **Respondent grew up in France** — Binary indicator derived from the question on country of birth and childhood.

Parental demographics (father and mother separately)

- **Born or raised in France** — Categorical indicator.
- **Year of arrival in France** — Categorical, with quartile-based cutoffs (father: before 2010, 2010–2014, 2015–2018, after 2018; mother: before 2014, 2014–2017, 2018–2020, after 2020).
- **Age** — Categorical, quartile-based.
- **Region of origin** — Based on UN sub-regional classification: North Africa, Sub-Saharan Africa, South and East Asia, Central and Western Asia, Latin America, Eastern Europe, Western Europe and USA, unknown origin.

Household characteristics

- **Education level** — Categorical: none, primary, lower secondary, upper secondary, higher education.
- **Number of children** — Categorical: 1, 2, 3 or more.
- **Family structure** — Categorical: both parents together, single mother, other arrangement.
- **Residency status** — Derived from *"What is your residency situation?"* Recoded into three categories: stable (French national or residence permit of 10 years or more), temporary (residence permit of less than 10 years), precarious (no permit, pending application, or other situation). "Don't know" and "prefer not to answer" were coded as missing.

C Main Estimates

Table C1: ITT estimates on parental language proficiency and engagement in schooling

Outcome	No covariates	With covariates	Post-double LASSO
Father's French level	0.20*	0.02	0.03
	(0.10)	(0.08)	(0.08)
Mother's French level	-0.05	-0.05	-0.06
	(0.12)	(0.06)	(0.05)
Linguistic fluency	0.07	-0.04	-0.04
	(0.10)	(0.07)	(0.07)
School activity participation	0.22**	0.28***	0.30***
	(0.10)	(0.10)	(0.11)
Home activities	-0.04	0.02	0.04
	(0.09)	(0.10)	(0.10)
Involvement in schooling	-0.05	-0.00	0.01
	(0.06)	(0.07)	(0.08)
Covariates	No	Yes†	BCH-DL‡
Obs. range	632-713	588-657	588-657

Notes: This table reports intention-to-treat (ITT) effects of OEPRE school availability on primary outcomes under three specifications. Column (1) uses no additional controls; column (2)^{dagger} includes the controls retained by iterative VIF selection (threshold 5 within each block, 10 on the combined set): Father born/raised in FR, Father year of arrival, Father age group, Father region of origin, Mother born/raised in FR, Mother year of arrival, Mother age group, Mother region of origin, Education level, Number of children, Family structure, Residency status, French nationality, Language at arrival (std.), Respondent is mother; school level: UPE2A, ULIS, total pupils, share of girls, teacher age, teacher seniority, teaching staff (FTE), share tenured, share of female teachers, QPV, educational city, REP, REP+, rural, intermediate urban, dense urban, log municipal population, median income, PCS advantaged, PCS intermediate, fragile in French, fragile in maths, satisfactory in maths; column (3)^{ddagger} uses covariates selected by double LASSO (BCH procedure, union of the outcome and treatment equations; selection engine: glmnet::cv.glmnet (validation croisee)) from the full pooled set of individual and school-level candidate covariates, with no prior VIF screening. All outcomes are standardised on the control-group distribution (mean 0, sd 1). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Standard errors are clustered at the school level. Sources: parents' survey and DEPP data [ARCHIPEL, 2019-2025].

Table C2: ITT estimates on parental civic, social, and economic outcomes

Outcome	No covariates	With covariates	Post-double LASSO
Civic engagement	0.10	0.01	0.06
	(0.11)	(0.10)	(0.10)
Trust in institutions	-0.08	-0.06	-0.05
	(0.08)	(0.09)	(0.09)
Sense of belonging	0.10	-0.02	0.03
	(0.10)	(0.10)	(0.09)
Gender attitudes	-0.11	0.04	0.03
	(0.09)	(0.10)	(0.10)
Self-esteem	0.08	0.20**	0.16
	(0.08)	(0.10)	(0.10)
Social network	-0.01	0.07	0.08
	(0.09)	(0.10)	(0.10)
Father's living conditions	-0.08	-0.11	-0.07
	(0.10)	(0.10)	(0.10)
Mother's living conditions	-0.19	-0.08	-0.07
	(0.15)	(0.09)	(0.09)
Covariates	No	Yes[†]	BCH-DL[‡]
Obs. range	708-713	653-657	653-657

Notes: This table reports intention-to-treat (ITT) effects of OEPRE school availability on secondary outcomes under three specifications. Column (1) uses no additional controls; column (2)^{dagger} includes the controls retained by iterative VIF selection (threshold 5 within each block, 10 on the combined set): Father born/raised in FR, Father year of arrival, Father age group, Father region of origin, Mother born/raised in FR, Mother year of arrival, Mother age group, Mother region of origin, Education level, Number of children, Family structure, Residency status, French nationality, Language at arrival (std.), Respondent is mother; school level: UPE2A, ULIS, total pupils, share of girls, teacher age, teacher seniority, teaching staff (FTE), share tenured, share of female teachers, QPV, educational city, REP, REP+, rural, intermediate urban, dense urban, log municipal population, median income, PCS advantaged, PCS intermediate, fragile in French, fragile in maths, satisfactory in maths; column (3)^{ddagger} uses covariates selected by double LASSO (BCH procedure, union of the outcome and treatment equations; selection engine: glmnet::cv.glmnet (validation croisee)) from the full pooled set of individual and school-level candidate covariates, with no prior VIF screening. All outcomes are standardised on the control-group distribution (mean 0, sd 1). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Standard errors are clustered at the school level. Sources: parents' survey and DEPP data [ARCHIPEL, 2019-2025].

Table C3: ITT estimates on children's test scores, with school covariates

	French test score			Mathematics test score		
	(1)	(2)	(3)	(1)	(2)	(3)
OEPRE school availability	0.11	0.10	0.12	-0.04	-0.01	0.01
	(0.11)	(0.11)	(0.10)	(0.10)	(0.09)	(0.09)
Observations	496	458	458	499	461	461
Covariates	No	Yes	BCH-DL	No	Yes	BCH-DL

Notes: This table reports intention-to-treat effects of OEPRE school availability on children's test scores. The outcomes are the child's most recent standardised score in French and in Mathematics (first children; z-scores on the control-school distribution of the same grade, evaluation years pooled within grade). For children of OEPRE participants, the most recent score is required to be strictly posterior to the household's OEPRE starting year; the few children of participants with no such score are excluded. Column (1) uses no additional controls. Column (2) includes the controls retained by iterative VIF selection (threshold 5 within each block, 10 on the combined set): father born/raised in France, father's arrival year, father's age, father's region of birth, mother born/raised in France, mother's arrival year, mother's age, mother's region of birth, education level, number of children, family organisation, residency status, French nationality, language score at arrival, respondent gender; school level: UPE2A, ULIS, total pupils, share of girls, teacher age, teacher seniority, teaching staff (FTE), share tenured, share of female teachers, QPV, educational city, REP, REP+, rural, intermediate urban, dense urban, log municipal population, median income, PCS advantaged, PCS intermediate, fragile in French, fragile in maths, satisfactory in maths. Column (3) uses covariates selected by double LASSO (BCH procedure, union of the outcome and treatment equations; selection engine: glmnet::cv.glmnet (cross-validated penalty)) from the full pooled set of individual and school-level candidate covariates, with no prior VIF screening. French test score: 54 terms retained, of which 24 school-level. Mathematics test score: 46 terms retained, of which 22 school-level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Standard errors are clustered at the school level. Sources: parents' survey and DEPP data [ARCHIPEL, EVA6, and REPÈRES, 2019-2025].

Table C4: ATT estimates on parental language proficiency and engagement in schooling

Outcome	(IV)	(CBPS)	(Entr. Bal.)
Father's French level	0.03	0.20*	0.20*
	(0.16)	(0.11)	(0.11)
Mother's French level	-0.10	-0.01	-0.01
	(0.12)	(0.08)	(0.08)
Linguistic fluency	-0.07	0.05	0.05
	(0.15)	(0.10)	(0.10)
School activity participation	0.57***	0.43***	0.43***
	(0.20)	(0.12)	(0.12)
Home activities	0.04	0.07	0.07
	(0.20)	(0.13)	(0.13)
Involvement in schooling	-0.00	0.05	0.05
	(0.15)	(0.09)	(0.09)
Covariates	Yes†	Yes‡	Yes‡
Obs. range	588-657	387-426	387-426

Notes: This table reports estimates of the effect of OEPRE participation on primary outcomes under three identification strategies. (IV)^{dagger} reports 2SLS estimates on the full sample, identifying the LATE on compliers with school-level treatment assignment as instrument; controls are the controls retained by iterative VIF selection (threshold 5 within each block, 10 on the combined set): Father born/raised in FR, Father year of arrival, Father age group, Father region of origin, Mother born/raised in FR, Mother year of arrival, Mother age group, Mother region of origin, Education level, Number of children, Family structure, Residency status, French nationality, Language at arrival (std.), Respondent is mother; school level: UPE2A, ULIS, total pupils, share of girls, teacher age, teacher seniority, teaching staff (FTE), share tenured, share of female teachers, QPV, educational city, REP, REP+, rural, intermediate urban, dense urban, log municipal population, median income, PCS advantaged, PCS intermediate, fragile in French, fragile in maths, satisfactory in maths. (CBPS)^{ddagger} and (Entr. Bal.)^{ddagger} report doubly-robust ATT estimates: weights balance the household matching covariates (respondent's gender (mother/father), French nationality, language score at arrival, respondent's age, mother's presence in France, region of birth (7 origin dummies, ref: Western Europe/USA), respondent's education level (4 dummies, ref: tertiary or above), number of children, family structure (parents together, single mother), residency status (temporary permit, precarious status; ref: stable)), and the weighted regressions condition on these same matching covariates plus the school-level controls selected by iterative VIF. School-level variables are excluded from the weighting model because entropy balancing fails to converge when they are added. The weighted estimators compare OEPRE participants from treatment schools with non-participating respondents from control schools; participants enrolled in control schools are excluded. All outcomes are standardised on the control-group distribution (mean 0, sd 1). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Standard errors are clustered at the school level. Sources: parents' survey and DEPP data [ARCHIPEL, 2019-2025].

Table C5: ATT estimates on parental civic, social, and economic outcomes

Outcome	(IV)	(CBPS)	(Entr. Bal.)
Civic engagement	0.02	0.13	0.13
	(0.21)	(0.12)	(0.12)
Trust in institutions	-0.12	0.01	0.01
	(0.18)	(0.11)	(0.11)
Sense of belonging	-0.03	-0.02	-0.02
	(0.20)	(0.13)	(0.13)
Gender attitudes	0.08	0.02	0.02
	(0.19)	(0.10)	(0.10)
Self-esteem	0.42**	0.20*	0.20*
	(0.21)	(0.11)	(0.11)
Social network	0.14	0.09	0.09
	(0.19)	(0.12)	(0.12)
Father's living conditions	-0.22	-0.08	-0.08
	(0.21)	(0.11)	(0.11)
Mother's living conditions	-0.15	-0.05	-0.05
	(0.19)	(0.10)	(0.10)
Covariates	Yes†	Yes‡	Yes‡
Obs. range	653-657	423-426	423-426

Notes: This table reports estimates of the effect of OEPRE participation on secondary outcomes under three identification strategies. (IV)^{dagger} reports 2SLS estimates on the full sample, identifying the LATE on compliers with school-level treatment assignment as instrument; controls are the controls retained by iterative VIF selection (threshold 5 within each block, 10 on the combined set): Father born/raised in FR, Father year of arrival, Father age group, Father region of origin, Mother born/raised in FR, Mother year of arrival, Mother age group, Mother region of origin, Education level, Number of children, Family structure, Residency status, French nationality, Language at arrival (std.), Respondent is mother; school level: UPE2A, ULIS, total pupils, share of girls, teacher age, teacher seniority, teaching staff (FTE), share tenured, share of female teachers, QPV, educational city, REP, REP+, rural, intermediate urban, dense urban, log municipal population, median income, PCS advantaged, PCS intermediate, fragile in French, fragile in maths, satisfactory in maths. (CBPS)^{ddagger} and (Entr. Bal.)^{ddagger} report doubly-robust ATT estimates: weights balance the household matching covariates (respondent's gender (mother/father), French nationality, language score at arrival, respondent's age, mother's presence in France, region of birth (7 origin dummies, ref: Western Europe/USA), respondent's education level (4 dummies, ref: tertiary or above), number of children, family structure (parents together, single mother), residency status (temporary permit, precarious status; ref: stable)), and the weighted regressions condition on these same matching covariates plus the school-level controls selected by iterative VIF. School-level variables are excluded from the weighting model because entropy balancing fails to converge when they are added. The weighted estimators compare OEPRE participants from treatment schools with non-participating respondents from control schools; participants enrolled in control schools are excluded. All outcomes are standardised on the control-group distribution (mean 0, sd 1). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Standard errors are clustered at the school level. Sources: parents' survey and DEPP data [ARCHIPEL, 2019-2025].

Table C6: Effects of OEPRE participation on children's test scores, with school covariates

	French test score			Mathematics test score		
	(IV)	(CBPS)	(Entr. Bal.)	(IV)	(CBPS)	(Entr. Bal.)
OEPRE participation	0.25	0.30*	0.30*	-0.03	0.04	0.03
	(0.25)	(0.16)	(0.16)	(0.22)	(0.12)	(0.12)
Observations	458	285	285	461	290	290
Covariates	Pre-det. + school	Matching set + school	Matching set + school	Pre-det. + school	Matching set + school	Matching set + school

Notes: This table reports estimates of the effect of OEPRE participation on children's test scores under three methods. The treated group consists of children of OEPRE participants from treatment schools; the comparison group consists of children of non-participating households from control schools. (IV) reports 2SLS estimates; the instrument is school-level treatment assignment and the coefficient identifies the LATE on compliers; the 2SLS conditions on the controls retained by iterative VIF selection (threshold 5 within each block, 10 on the combined set): father born/raised in France, father's arrival year, father's age, father's region of birth, mother born/raised in France, mother's arrival year, mother's age, mother's region of birth, education level, number of children, family organisation, residency status, French nationality, language score at arrival, respondent gender; school level: UPE2A, ULIS, total pupils, share of girls, teacher age, teacher seniority, teaching staff (FTE), share tenured, share of female teachers, QPV, educational city, REP, REP+, rural, intermediate urban, dense urban, log municipal population, median income, PCS advantaged, PCS intermediate, fragile in French, fragile in maths, satisfactory in maths. (CBPS) and (Entr. Bal.) report ATT estimates from doubly-robust weighted OLS using CBPS and entropy balancing weights respectively; the weights balance the household matching covariates (an indicator for the respondent being the mother, French nationality, the standardised language score at arrival, the respondent's age, whether the mother was born or raised in France, the mother's region of birth, the respondent's education level, the total number of children, family organisation and residency status), the weighted regressions condition on these same matching covariates plus the school-level controls listed above, and school-level variables are excluded from the weighting model because entropy balancing fails to converge when they are added. The Observations row gives each model's estimation sample. Outcomes are the child's most recent standardised score (first children; z-scores on the control-school distribution of the same grade); for children of OEPRE participants, the most recent score is required to be strictly posterior to the household's OEPRE starting year, and the few children of participants with no such score are excluded. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Standard errors are clustered at the school level. Sources: parents' survey and DEPP data [ARCHIPEL, EVA6, and REPÈRES, 2019-2025].

Chapter 2

Educate and Punish: Allocation, Displacement, and the Consequences of an Intermediate Juvenile Sanction

Evidence from French Closed Educational Centres

Abstract

Only a small fraction of the minors a justice system processes reaches detention, yet almost all the causal evidence is focused on it. The effects of measures that precede it, including those meant to replace it, are largely undocumented, partly because their sign is not obvious and depends on what they displace: prison for some minors, community supervision for others. I study this question by focusing on the French closed educational centre (CEF), a residential measure created in 2002 as an alternative to incarceration, using administrative records that report the placement an educator recommends to the judge before the hearing. That recommendation captures what the registers omit about the minor's family, housing and schooling situations, and predicts placement far better than the recorded judicial trajectory. Minors recommended for a centre and sent elsewhere overwhelmingly receive community supervision, fewer than one in ten detention suggesting CEF placement adds a residential component to a supervision order rather than replacing custody. Against that comparison group, placement raises detention within twelve months of the decision by 8 percentage points, from a counterfactual rate of 44%. The effect is largest against community supervision (21 points) and remains positive against detention (11 points), the response the centre was created to avoid. The placement is attached to a supervision order that a breach of the centre's rules can revoke, sending the minor to prison: a quarter of the detentions within a year begin as the measure itself ends, against 7% after a supervision order alone. What sends these minors to prison is therefore partly the legal form of the sanction, not what the programme delivers.

JEL codes: K42, K14, J13, I28, J24
Keywords: Juvenile delinquency, Custodial alternatives, Desistance, Rehabilitation programmes, Criminal capital

I am grateful to the French Ministry of Justice (Service de la Statistique, des Études et de la Recherche, SSER and Service des Études, de la Recherche et des Évaluations, SEREV-PJJ) for making the data available, especially Pascal Chevalier, Malik Koubi, and Alice Simon for their administrative and logistical help, as well as Sylvain Renault and Isabelle Thing-Leoh for their help in understanding the data. I thank participants of the DIAL Seminar (2022) LIEN Seminar (2026), and SDL Dauphine (2026). This project benefited from very helpful feedback and suggestions from Roberto Galbiati, Flore Gubert, Élise Huillery, and Liam Wren-Lewis. I am particularly indebted to Arnaud Philippe for his support and help right from the start of this project.

Contents

1 Introduction . . . **73**
2 Institutional Setting . . . **78**
2.1 The Closed Educational Centre . . . 78
2.2 The Other Responses Available to the Court . . . 80
2.3 Allocation of Minors to a Closed Educational Centre . . . 81
3 Data . . . **82**
3.1 Data Sources and Sample Selection . . . 82
3.2 Variables of Interest . . . 85
3.3 Descriptive Statistics . . . 86
4 Empirical Strategy . . . **88**
4.1 Identification Challenge . . . 88
4.2 Defining the counterfactual . . . 89
4.3 Documenting the allocation to a closed educational centre . . . 91
4.4 Estimating the effects of placement in a closed educational centre . . . 92
5 Results . . . **95**
5.1 The decision process of the placement . . . 95
5.2 The effects of the placement . . . 99
6 Mechanisms . . . **110**
7 Conclusion . . . **116**
References . . . **118**
Appendix . . . **121**
A The French Juvenile Justice System . . . **121**
A.1 Judicial Process and the Severity Ladder of Sanctions . . . 121
A.2 Judicial Measures Available to a Juvenile Court . . . 124
A.3 Geographic Roll-out of the Network . . . 127
B Data Coverage and Measurement . . . **130**
B.1 Prescriber's Department Coverage (2004-2006) . . . 130
B.2 Detention: Data Availability (2010-2012) . . . 133
B.3 Variable Definitions . . . 134
B.4 Baseline Characteristics of the Analytical Cohorts . . . 136
B.5 Selection into the Report-Linked Sample . . . 136
C Comparison Group, Balance and Sensitivity . . . **139**
D The Placement Decision . . . **150**
E Additional Estimates . . . **156**
F Estimates Anchored at the End of the Measure . . . **159**
G Additional Mechanism Evidence . . . **163**

1 Introduction

Contact with the justice system during adolescence coincides with the years in which schooling, training and early labour-market attachment are decided, and the existing evidence points to durable consequences: juvenile incarceration and and pre-trial detention both reduce high-school completion, while also increasing later involvement with criminal justice through adult conviction and arrest by age 19 (Aizer and Doyle Jr., 2015; Baron et al., 2023). Juvenile delinquency has nonetheless received less causal attention than adult crime, and the existing evidence has mainly concentrated on detention, even though it is intended to be exceptional in juvenile justice. The Convention of the Rights of the Child requires detention or imprisonment to be used only as a measure of last resort and for the shortest appropriate period (United Nations Children's Fund, 2021b). Yet United Nations Children's Fund (2025) estimates that 259,000 children aged 5 to 17 were in detention worldwide in 2024[1]. What justice systems do before detention remains largely undocumented.

Juvenile justice systems are built around a tension that adult systems do not face in the same terms: the same decision is expected to sanction the offence and to give particular weight to education as a rehabilitation mechanism. The recommendations of international organisations reflect this by placing particular emphasis on care, guidance, supervision orders, foster care, residential communities, and other educational settings as alternatives to imprisonment (United Nations Children's Fund, 2021a). Several European systems answer this tension by developing residential measures that combines education with different levels of constraint outside ordinary prison, such as public youth-protection institutions in Belgium (Fédération Wallonie-Bruxelles, 2026), the *collocamento in comunità* in Italy (Ministero della Giustizia, 2026), or the Youth Rehabilitation Orders with Intensive Supervision and Surveillance in England and Wales (Youth Justice Board for England and Wales, 2024). Despite important institutional differences, whether these measures reduce or increase coercion depends on what alternative they displace. A residential measure can reduce coercion for a minor who would otherwise have been detained, or increase it for one who would otherwise have remained under community supervision.

Whether such a measure actually protects minors from prison or brings them closer to it therefore depends on which of the two populations it actually receives. In this case, the way minors are selected into the measure is an important part of what the policy does, because the same intervention has different implications depending on what it replaces, and its effect does not rely solely on its contents.

Estimating the effect of such a measure is difficult for reasons specific to its position in the system. Minors who receive it are selected on a judicial trajectory that also predicts the outcomes of interest, so a naive comparison is uninformative. Identification strategies used in other settings to address this selection problem rely on features not available in the French registers: there is no magistrate identifier, no capacity saturation of the residential facilities, and no available instrument with a credible exclusion restriction (See Section 4.1 for a discussion). Because the measure sits between incarceration and community supervision rather than at the top of the

[1] Global administrative data do not provide a comparable count of all children who come into conflict with the law, so this figure cannot be converted into a worldwide share of children involved with justice.

scale, its effect also has no obvious direction. It has to be estimated against the specific measures it displaces, while the counterfactual measure for a minor placed is not observed. What the question needs is direct information on the assessment that precedes the decision of the judge: attached to each case is a social-educational report (*Recueil de Renseignements Socio-Éducatifs*, RRSE) written by the duty educator before the hearing, which records the placement the educator recommends to the judge. That recommendation summarises information observed by the educator but unavailable in the administrative record such as the family situation, the living conditions, schooling or training, any child-protection history, and the state of the minor and the parents at the time of the interview. Comparing minors who received the same recommendation therefore allows to control for an important summary of information that would otherwise be unobserved in the administrative data.

In order to investigate these questions, I use the French context to study one such measure: the closed educational centre (*Centre Éducatif Fermé*, CEF), created by the 2002 *Perben I* law. In France, roughly 130,000 minors receive a penal response each year; a few thousand are incarcerated and around 1,000 enter a closed educational centre. The centre hosts 12 minors for a 6 months programme, with a staffing target of about 2 adults per minor, internal schooling delivered by seconded teachers, and vocational training (IGSJ and IGAS, 2015; Direction de la Protection Judiciaire de la Jeunesse, 2016). Its daily cost per minor reaches 705 € in the public sector, roughly twice that of an ordinary residential placement (Cour des comptes, 2023). What mainly separates it from incarceration is the legal format: the placement cannot be ordered on its own, only as an accessory to a coercive order, in 92% of cases a judicial supervision order (Direction de la Protection Judiciaire de la Jeunesse, 2021). Breaching the rules of the centre allows that order to be revoked, and the revocation sends the minor to prison. The centre is therefore a rehabilitative programme delivered under a suspended threat of custody, and has been conceived politically as the last step before incarceration.

The effect of such a placement on later offending is not obvious in theory. The juvenile evidence on custody does not settle the sign, as incarceration lowers reconviction in one setting (Hjalmarsson, 2009), raises it under a harsher custodial regime (Lotti, 2022), and has different effects across offence types (Eren and Mocan, 2021). Despite the scarce evidence, structured programmes delivered to people already in the system can work, whether inside a juvenile detention centre, in prison or prison-wide (Heller et al., 2017; Arbour, 2022; Totarelli, 2024), and the institution setting matters too, since conditions, facility quality and the orientation of the regime all shape post-release behaviour (Drago et al., 2011; Mastrobuoni and Terlizzese, 2022; Tobón, 2022; Bhuller et al., 2025). Evidence comparing community sentence with custody is also mixed, as incarceration raises later offending in one system (Michel and Hémet, 2022) and lowers it in a system organised around rehabilitation (Bhuller et al., 2020). Additionally, placing 12 minors with judicial records in one facility for several months exposes each of them to the others, and peer composition inside custodial institutions has been found to raise subsequent offending (Bayer et al., 2009; Stevenson, 2017; Ouss and Philippe, 2025; Drago and Galbiati, 2012).

A last channel runs through the legal instrument rather than through the actual experience of

placement. Supervision regimes can generate custody through the violation of their conditions. On probation, the custody produced by this mechanism is only partly targeted (Franco et al., 2024; Rose, 2021), while on parole, the random assignment of board members raises returns to prison through breaches of conditions while lowering convictions for new crimes (Arbour and Marchand, 2025). Part of any measured effect on later detention may therefore be produced by the judicial order that carries the placement rather than by the placement itself.

Which channel dominates depends on who is placed, so the selection of minors into the measure is the first question I address. Judicial decision-making is itself a well studied object (Depew et al., 2017; Harrington et al., 2026; Stevenson and Doleac, 2024; Angelova et al., 2026) but these designs do not necessarily describe the population a given measure ends up capturing. However, this determines whether a measure created as an alternative to incarceration functions as one.

I use two administrative sources derived from the information systems of the French juvenile justice directorate (*Protection Judiciaire de la Jeunesse*, PJJ). PARCOURS covers 2021 to 2024 and links 80,201 penal measures to a report, of which 4,302 carry a recommendation for a closed educational centre. On a subsample it also records the nature of the offence, and it observes family and demographic characteristics older registers do not. On this source, the comparison group is drawn from the minors the institution judged suitable for the measure, namely those recommended for a closed educational centre and sent elsewhere. The recommendation enters as a control, so that selection on what the educator observed can be tested directly. Its sample and its horizon are short, so I turn to the historical registers of the same directorate, GAME/IMAGES, from 2012 to 2022 and 934,374 spells. They record neither the recommendation nor the offence, so the comparison group there must be constructed: I estimate a propensity score on 22 pre-placement covariates, restrict the sample to the region of common support, and reweight untreated spells by entropy balancing. I read those estimates as complementary evidence under a weaker conditioning set, not as bounds in any formal sense. Interestingly, the registers also allow a much finer reconstruction of the judicial trajectory, and carry a monthly sub-base that records what happens inside the placement.

In order to first ask what governs the decision, I estimate a multinomial logit of the response a minor receives among community supervision, residential placement, closed educational centre and detention, on the trajectory observed before the decision. I measure how much each block of information adds to the explanation. The recorded judicial history predicts well who will later be detained but predicts a centre placement poorly: the area under the curve for the centre is 0.79 on the trajectory alone and reaches 0.88 once the educator's recommendation is added, while on detention the order is reversed and the nature of the offence contributes more than the recommendation does. Detention is decided on what the minor did and the closed centre on an assessment of what the minor needs, and that assessment is what the administrative trajectory fails to reproduce. Geography adds little once the trajectory and the recommendation are held constant, which fits evidence that judicial practices are local and learned rather than determined by the case at hand (Abrams et al., 2022). What remains unexplained is governed less by the minor than by the receiving side: minors are frequently placed outside the prescribing

department, facilities differ in the severity of the profiles they admit from the same courts in the same years, and admissions arrive in batches, so that who lives with whom is settled by judicial batches rather than by the timing of individual trajectories.

I then ask who this allocation captures in practice, considering that the measure is designed for repeat offenders as a "last chance" before prison. Among minors whose report recommended a closed centre and who were not placed in one, the large majority received community supervision, a small share another residential placement, and fewer than 1 in 10 detention. The reweighted comparison group on the historical registers is composed of the same families in the same proportions, which is reassuring given that the two counterfactuals are built in entirely different ways. Most of that community supervision is the judicial supervision order itself, the very vehicle to which the placement is attached. For most of the minors it receives, the closed educational centre therefore operates as a residential addition to a supervision order they would have received anyway rather than as a substitute for custody, so the estimates measure what happens when a minor is moved one step further into the system and not what happens when a minor is spared prison. That distinction matters, because being processed more heavily carries a cost of its own, distinct from the sanction that follows (Agan et al., 2023; Mueller-Smith and Schnepel, 2021), even if both those studies concern indivudals with lighter records than those studied here.

Turning to effects, I find that placement in a closed educational centre raises the probability of a subsequent detention episode. Outcomes run over fixed windows from the decision date, so the two arms are observed over the same length of time and a custody that revokes the placement while it is still running counts as one of its consequences. Within the recommended population, where the comparison group is made of minors judged suitable for the measure and sent elsewhere, the probability of a detention episode within 12 months of the decision rises by 8.0 percentage points on a counterfactual rate of 43.5% when the recommendation is the educator's narrative, and by 9.7 points when it is the formal post-investigation decision. At 18 months neither estimate are significant but are estimated on smaller samples and thus, less precise. On the historical registers, which are large enough to follow the same window on a much wider population, placement increases the probability of a detention episode within 12 months by 20.4 points on a counterfactual rate of 21.6%, and 21.5 points at 18 months. Estimated against each displaced alternative separately, the effect runs from 21.5 points against community supervision down to 11.0 points against detention, suggesting that it is largest exactly where the measure is most used, and is non-null even against the response the centre was created to avoid.

I then turn to the mechanisms underlying this effect by first measuring, among the detentions following within 12 months, how often the custody ends the placement measure, which is what a revocation produces. That share reaches 26.8% after a closed centre placement, against 16.8% after an ordinary residential placement, 11.3% after a reinforced centre and 7.1% after a judicial supervision order alone. Removing those episodes, by measuring outcomes from the end of the measure rather than from the decision, lowers the twelve-month estimate on the registers from 20.4 to 13.6 points, which is one reading of what this channel would indicate. Using monthly housing records allows me to explore what happens inside the residential placements.

I find that incidents and recorded absences are more frequent inside a closed centre than inside the reweighted residential comparison, and the adverse pattern is concentrated on stays cut short before the final 2-month phase that prepares the exit of the minor. The relevance of the revocation channel for identification is that the device attaches to the measure and not to the minor: whatever unobserved severity placed minors carry, a comparison minor under the same supervision order faces no residential rules to breach and no facility to run from, so a part of the estimated gap is produced by the instrument and would survive even if the two arms were identical in every unobserved respect.

When compared with the intention of the 2002 law, these results describe a displacement rather than a failure of the programme. The measure was designed as the last step before prison and is delivered mainly to minors with judicial supervision, and under the very order that makes the revocation possible. The design rests on selection on observables, and I run three exercises to quantify how strong unobserved selection would have to be to overturn it. Holding the educator's recommendation constant reduces the 12 months estimate from 17.0 to 14.0 percentage points, indicating that information observed by the educator but absent from the administrative trajectory accounts for part, but not most, of the estimated gap. Following Oster (2019)'s approach, selection on unobservables would have to be four times as strong as selection on the observed covariates to drive the estimate on the registers to zero. On the recommended population, which carries the main estimate, the same exercise indicates that the coefficient moves *away* from zero as the covariates are added, so proportional selection on what is not recorded would amplify the effect rather than explain it away. The revocation device operates on the placed arm alone, so part of the measured gap is produced by this rather than by any difference between the minors. What the results do not establish is whether the programme itself is effective or not; the estimate is smallest on the population the institution itself selected for the measure, although the design cannot say whether this gap is heterogeneity or the stricter conditioning imposed by the sample of CEF-recommended minor.

The paper makes four contributions. It treats the allocation of an intermediate measure as an object of study rather than as a source of variation, and examines what determines the decision of placement when the we are able to observe the institution's own pre-decision assessment together with administrative record, rather than using judicial assignment mainly as a source of variation to identify the effects. It documents the alternatives that the CEF placement displaces in two independently constructed data sources, allowing each alternative to be examined separately rather than only pooling them in a single comparison group. It shows that the revocation mechanism already documented on community and post-release supervision also operates on a measure the law presents as an alternative to incarceration, and gives evidence consistent with it accounting for an important part of the estimated effect. It estimates the effect of a measure located between community supervision and prison, where the juvenile literature has concentrated on the custodial margin and remains divided, while the intermediate measures that carry most of the response to serious juvenile offending recieved much less attention.

Finally, the analysis does not open the black box of the closed centre. Peer composition, which the literature identifies as a first-order channel, cannot be isolated here, and the same

holds for the educational, vocational and health components of the programme, which may well work for some minors without the data allowing me to say for whom or why. Given how much of the intervention is devoted to human-capital formation, the outcomes of interest are not only judicial but also schooling, training, employment and health, which the present source does not observe. What works inside places of deprivation of liberty, and whether a programme of this intensity fails because of who it receives or because of how it is delivered, remains open.

The paper proceeds as follows. Section 2 describes the French juvenile justice system, the position of the closed educational centre among the responses available to a court, and how minors are allocated to one. Section 3 presents the two administrative sources and the sample. Section 4 states the identification problem, establishes which measures the placement displaces, and presents the estimation. Section 5 reports the determinants of the placement, and its effects on later judicial trajectories against each alternative it displaces. Section 6 turns to the channels, in particular the revocation of the order that carries the placement and events happening during the placement. Section 7 concludes.

2 Institutional Setting

This section describes the closed educational centre, the placement this paper studies (Section 2.1), the other responses among which a court chooses (Section 2.2), and the way minors are allocated to a centre in practice (Section 2.3).

2.1 The Closed Educational Centre

Created by the 2002 *Perben I* law, the CEF is the first juvenile institution since the founding 1945 *Ordonnance* to incorporate a custodial logic onto the educational placement framework. Legally, CEF placement is not a sentence: it is a measure that can be ordered only as an *accessory* to a coercive judicial vehicle, namely judicial supervision (*contrôle judiciaire*, CJ), suspended sentence with probation (*sursis avec mise à l'épreuve*, SME), conditional release, or external placement (IGSJ and IGAS, 2015; Direction de la Protection Judiciaire de la Jeunesse, 2016). The placement therefore retains the doctrinal status of an educational measure, yet is delivered inside a facility that is operationally a closed institution; its "closed" character is legal rather than architectural, the operational containment lying in the threat of incarceration following any breach of the placement rules (Vuattoux, 2016). The CEF is, in that sense, a hybrid programme that imports a carceral logic into a system organised around education for the first time since 1945, and that the sociological literature has read as a hardening of the post-war doctrine of educational primacy (Mucchielli, 2005; Carlos, 2020). Empirically, the dependence on a coercive vehicle is overwhelming: 92% of minors placed in a CEF in mid-2021 were under judicial supervision, 7% under a suspended sentence with probation, and 1% under a sentence adjustment (Direction de la Protection Judiciaire de la Jeunesse, 2021). A breach of the placement obligation triggers a revocation of the underlying vehicle, which may then be converted into pre-trial detention or imprisonment, a mechanism documented institutionally (Inspection générale de la justice, 2025) and central to the empirical analysis of Section 6. The closed

educational centre is the most resource-intensive non-custodial modality of the system. Each closed centre is statutorily organised around 12 places, with a staffing target of 26.5 full-time-equivalent posts per facility, including 17 specialised educators, three management and clinical posts (director, head of educational unit, psychologist), 1.5 health-care posts, and technical instructors running at least three on-site vocational workshops. Schooling is delivered internally by teachers detached from the national education system, and the medical component relies on partnerships with the regional health agency and with addiction-treatment centres (Direction de la Protection Judiciaire de la Jeunesse, 2016; IGSJ and IGAS, 2015). The programme is designed in three sequential phases of approximately two months each: (i) an arrival and adaptation phase, (ii) a structured educational and vocational work phase, and (iii) an exit-preparation phase that anticipates the minor's return to the family or to a community-based educational setting. A daily cost per minor of 705 € in the public sector and 571 € in the private associative sector in 2019 makes the closed centre the most expensive non-custodial placement of the system, roughly twice the cost of the non-specialised foyer (Cour des comptes, 2023). The structures rolled out gradually after 2002, in three identifiable phases presented in Figure 2.1: a first wave between 2003 and 2010 in rural and peri-urban areas, an extension phase between 2011 and 2017 in denser urban areas, and a new-generation phase initiated in 2018.

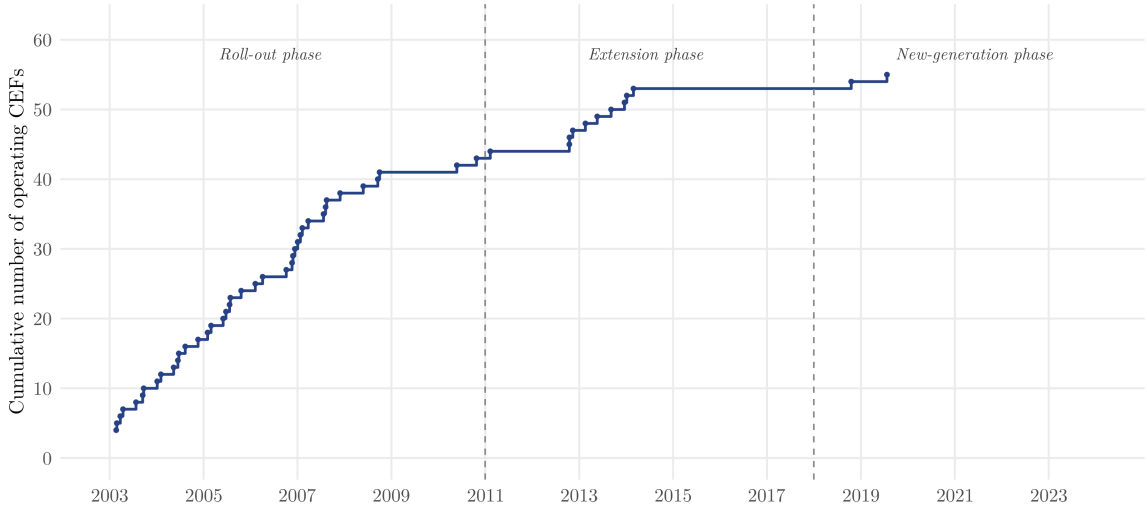

Figure 2.1: Cumulative roll-out of CEFs in France, 2003-2022.

Notes: The figure plots the cumulative number of operating closed educational centres as a function of the opening date, drawn from the DPJJ registry. The three policy phases are demarcated by dashed vertical lines: roll-out (2003-2010), extension (2011-2017), and new generation (2018 onward). Source: DPJJ registry of accredited closed educational centres.

Authorised and operational capacity grew alongside the network and the national occupancy rate fluctuated around 75% across the period[2]. The growth of capacity during the period, the geography of the network at the end of the rollout, and the year-by-year geographic expansion across these three phases are reported in Figures A3-A5 of Appendix A.

[2]Occupancy is measured in different ways across sources and the figures are not comparable: counts of minors present give around two thirds of capacity, whereas occupied days over capacity-days computed from the administrative records exceed it, because a place stays assigned to a minor who has run away or been incarcerated, and because some places are frozen without being withdrawn from the authorised count.

Minors placed in a closed centre are heavily selected on prior judicial trajectory and on socio-educational vulnerability. In the most recent national snapshot, 86% were already under PJJ supervision before the indexed placement and only 13% had no prior judicial history; three out of four were aged sixteen or seventeen, 95% were boys, and 81% had at least one diagnosed addiction (Direction de la Protection Judiciaire de la Jeunesse, 2021). Sociologists studying the daily life of the structure document a mode of supervision in which the constant proximity of the threat of incarceration shapes the educational relation and inserts a carceral logic into the placement (Vuattoux, 2016; Farcy-Callon, 2016).
The place of the closed educational centre among the responses a juvenile court can order, the three stages of the penal pathway from offence to disposition, and the full list of measures with their target population, minimum age and operational setting, are presented in Appendix A.

2.2 The Other Responses Available to the Court

Community supervision. Most minors under a penal measure are followed in the community rather than placed. An educator meets the minor and the family, follows schooling and any health or training arrangement, and reports back to the judge, and the measure is reviewed and adjusted as the situation changes. Part of this follow-up can be ordered under a judicial supervision order (*contrôle judiciaire*, CJ), a coercive vehicle that attaches obligations to the minor, such as a curfew, a prohibition on meeting certain people, or a duty to attend school or training, and that the judge can revoke if those obligations are breached. A placement in a closed educational centre is not decided on its own: it is an obligation added to a judicial supervision order. Therefore, a minor left under the order alone is under the same coercive vehicle without the residential rules that come with the centre.

The PJJ operates five residential modalities into which a juvenile court can place a minor. I compare them in Table 2.1 along their year of creation, target population, capacity, length of stay, and daily cost per minor. The non-specialised end of the spectrum gathers individualised housing units, collective foyers, and family-host arrangements; the reinforced educational centre (CER) sits one step up; the closed educational centre (CEF) sits one step further up still; and detention closes the spectrum. The five modalities coexist in the system and a minor can transit between them within a single judicial trajectory; the first block of Table 2.1 reports the corresponding spell counts and durations observed in the administrative records over the 2012-2022 window.

From foyers to detention. Collective foyers (*Unités Éducatives d'Hébergement Collectif*) host eight to twelve minors under round-the-clock supervision in a non-secured group setting, and individualised housing units (*Unités Éducatives d'Hébergement Diversifié*) scatter minors across apartments and host families at lower supervision intensity; together they carry the majority of the placement volume, at a daily cost of 323 € per minor in 2021 (Cour des comptes, 2023). One step up, the reinforced educational centre (CER), created in 1996, organises a short rupture from the minor's usual environment, three to five months structured by intense outdoor activities, for up to eight minors (CNAPE, 2012); a breach of its internal rules does not mechanically convert

Table 2.1: Comparison of residential modalities for juvenile offenders in France

	CEF	CER	Other placement	Detention
From DPJJ administrative records (spells 2012-2022)				
Number of observed spells (2012-2022)	19,265	13,092	62,439	47,875
Mean duration (days)	127	97	133	92
Median duration (days)	133	99	92	58
Share male	94.9%	92.8%	91.0%	95.8%
Mean age at start (years)	16.2	16.4	16.5	16.9
From institutional sources				
Year of creation (legal vehicle)	2002 (Loi Perben I)	1996 (circulars)	Pre-existing (PJJ)	2002 (EPM, Loi Perben I) Pre-existing (QM)
Active facilities (2022)	53	∼64	Hundreds	6 EPMs 45 QMs
Average authorised capacity per facility	12 places	8 places	8-12 places	60 places (EPM) 8-12 places (QM)
Staffing intensity per facility	26.5 FTE / facility (∼2.2 staff per minor)	No national norm	Varies by structure	117 FTE / EPM (1.7 minors per educator); no facility-level norm in QMs (5 minors per educator)
Average daily cost per minor	705 € (SP, 2019) 643 € (SAH, 2021)	648 € (2021)	323 € (2021)	601 € (EPM, 2021) 144 € (QM, 2021)
Statutory target population	Repeat or grave offenders, 13-18	First or moderate offenders, 13-18	Any minor under educational order	Pre-trial or convicted, 13-18

Notes: This table compares the four main residential modalities for juvenile offenders during the 2012-2022 analytical window. Upper block: statistics computed from the unified DPJJ administrative records produced by the cleaning pipeline. A spell is a continuous period of taking-charge by a single institutional modality. CEF and CER spells are identified by the institutional event group field. The Other placement column pools collective foyers (UEHC), individualised housing units (UEHD), out-of-foyer educational placements, and family hosts. Detention spells are drawn from the GAME detention sub-base. Lower block: figures sourced from IGSJ and IGAS (2015), Cour des comptes (2023), and Inspection générale de la justice (2025). SP refers to the public sector and SAH (*secteur associatif habilité*) refers to the private associative facilities accredited and partially funded by the PJJ. EPM denotes the juvenile detention centre model created by the Loi Perben I of 2002; QM denotes juvenile units within standard adult prisons. Staffing intensity figures are statutory norms drawn from Direction de la Protection Judiciaire de la Jeunesse (2016) for the closed centre and from Cour des comptes (2023) for juvenile detention centres and juvenile units; no national facility-level staffing norm is published for reinforced centres or for the residual placement modalities. The active CEF count and the average authorised capacity per CEF are recomputed from the DPJJ registry for internal consistency. As of 1 September 2026, public-sector CEFs are scheduled to be transformed into Unites Judiciaires a Priorite Educative (Ministère de la Justice, 2026). Source: SSER data [GAME/IMAGES, 2012-2022], DPJJ registry of accredited facilities and published inspectorate reports.

the placement into custody, and its daily cost of 648 € approaches that of the closed centre over a substantially shorter stay. Detention, finally, is served either in the juvenile unit of an adult prison (*Quartier pour Mineurs*, QM), which houses roughly two thirds of detained minors at 144 € per day, or in one of the six purpose-built juvenile detention centres (*Établissement Pénitentiaire pour Mineurs*, EPM) created by the same 2002 statute as the closed centre, which host up to 60 minors in living units of around ten; the four-fold cost gap reflects supervision intensity rather than a different population, minors in the two settings being similar in age, length of stay and offence type.

2.3 Allocation of Minors to a Closed Educational Centre

A closed educational centre placement is overwhelmingly ordered by a juvenile judge: 70% of the placements observed in 2021 were ordered by a juvenile judge, 26% by an investigating judge, and 3% by a juvenile court (Direction de la Protection Judiciaire de la Jeunesse, 2021). The dominant institutional gateway is the *défèrement*, that is the fast-track presentation of the minor to a magistrate within twenty hours of the end of police custody (Direction de la Protection Judiciaire de la Jeunesse, 2020). Of the closed-centre placements observed in 2021, 59% were ordered immediately at the conclusion of a *défèrement* hearing and 41% in a planned admission; 78% of the minors arrived at the centre on the day the placement order was signed (Direction de la Protection Judiciaire de la Jeunesse, 2021). For the majority of placements, the decision

is therefore taken under time pressure.

Two institutional inputs shape that decision. The first is the recommendation of the duty PJJ educator at the court (*Mesure Éducative auprès du Tribunal*, MEAT), which produces a socio-educational diagnostic of the minor (*Recueil de Renseignements Socio-Éducatifs*, RRSE) and proposes an educational orientation, mandatorily before any pre-trial detention warrant can be issued. On a national survey of 360 deferments in 2019, the magistrate followed the duty educator's recommendation in roughly half the cases: 53% for a proposed penal placement, 64% for proposed judicial supervision (Direction de la Protection Judiciaire de la Jeunesse, 2020). When the duty educator's recommendation was a penal placement, the closed centre captured 47% of the proposals, the collective foyer 33%, and the reinforced centre 11%. Conditional on ordering a penal placement, the magistrate then validated the type of structure proposed almost systematically. The qualitative evidence collected by the same survey suggests that the institutional decision turns on the trajectory of the minor more than on the offence.

The second institutional input is the local availability of a closed educational centre. The duty educator does not propose a placement modality from a pure preference ordering; she contacts a series of facilities and proposes the modality that has an open slot. On the same 2019 survey, at the *défèrement* stage, 41% of all educator proposals are explicitly declared to be *"made by default"* for lack of a better-fitting available place. The median educator who recommended a penal placement contacted 5 establishments before settling on a proposal. The empirical consequence is that minors whose educator's proposal was made by default go to detention 1 time out of 2.5, against 1 out of 6.5 across the full panel (Direction de la Protection Judiciaire de la Jeunesse, 2020). The Inspection générale de la justice (2025) concurs in a 2025 report, noting that juvenile judges routinely deplore the lack of available closed-centre places and that the geographical distribution of the network is not aligned with the geographic distribution of juvenile delinquency. Geographical mismatch is also visible at the minor level: only 22% of the closed-centre placements observed in mid-2021 were in the minor's parental department, 30% were in another region, and 94% of the girls were placed out of region (Direction de la Protection Judiciaire de la Jeunesse, 2021), despite the official guidelines (up until 2016) that explicitly invoked geographical proximity as a criterion to maintain family ties during the stay and to support the exit phase of the programme (Inspection générale de la justice, 2025).

3 Data

3.1 Data Sources and Sample Selection

GAME and IMAGES. I use individual-level administrative records compiled by the French juvenile justice directorate (Direction de la Protection Judiciaire de la Jeunesse, hereafter DPJJ) for the period 2004 to 2022. Until 2021, the directorate ran two parallel information systems: GAME, which recorded measures and placements taking place in public-sector facilities directly operated by the DPJJ, and IMAGES, which recorded the equivalent activity for private associative facilities accredited and partially funded by the directorate. The two systems share the same structure and minor-level identifiers, and I treat them as a single source of measures

throughout. The unit of observation in this combined source is one prescribed judicial measure: each row gives the measure type, the prescribing court and the prescriber department, the dates of decision and of taking-charge, and a small set of demographic attributes on the minor (encrypted identifier, year of birth, sex, country of birth code). Detention episodes are recorded in a separate sub-base of GAME and are merged at the minor level downstream. The records describe each measure as an administrative object, not as a substantive intervention: the only proxy for the content of a measure is its family code and the legal vehicle that prescribes it. There is no description of the offence, no information on the minor's family situation or household income, no school or employment outcome, and no textual content from the case file. The minor's spatial trajectory is observed through the prescriber department of each measure as department of residence is not available; a separate database of facilities records the geographic location of each operating institution. The detention sub-base has very thin coverage before 2010 and the prescriber department field is partially missing in 2004 to 2006, both documented in Appendix B; these two coverage gaps drive the choice of the analytical window discussed below.

I reshape the row-level records into one row per spell, defined as a continuous period during which the same minor is taking charge of the same kind of measure under the same legal vehicle, in two parallel pipelines, one for the measures recorded in GAME and IMAGES and one for the detention episodes of the GAME sub-base. Spells still open at the end of 2022 are right-censored at that date.

Supplementary administrative records on placement events. A complementary administrative source derived from GAME and IMAGES decomposes each placement measure month by month and records, for every observed month, the number of days the minor spent in each housing modality of the placement (collective, individual, family-host, camp-style break) and the number of days lost to absence categories (permission, incarceration during the measure, hospitalisation, runaway, residual administrative interruptions). However, coverage is partial and uneven across placement types. Matching a placement spell to this source on the minor, the facility unit, the taking-charge identifier and the calendar month, which is the rule the mechanism analysis applies, returns at least one observable month for 41.8% of closed educational centre spells, 15.7% of reinforced centre spells and 65.5% of foyer and family spells indexed between 2012 and 2022, and for 53.8% of placement spells taken together.[3] The coverage is therefore neither exhaustive nor uniform, and it is selective on observables, but it is wide enough on the two arms the mechanism analysis compares, the closed centre and the residential placement, to support an exploratory analysis of what happens within the stay (Section 6).

PARCOURS. PARCOURS is the operational successor to GAME and IMAGES. It has been progressively rolled out from 2018 onwards and replaces the two parallel systems with a unified information system that covers, on a substantially shorter time window, every penal and civil

[3]Relaxing the unit and taking-charge conditions, and requiring only that the minor be observed in the source during the months the spell spans, raises these shares to 56.9, 36.8 and 75.7% respectively. Requiring only that the minor appear in the source at some point, at any date, raises them further to 72.4, 55.9 and 83.0%.

intervention prescribed by the directorate. The structure of PARCOURS is richer than the one of GAME/IMAGES along three dimensions that matter for the analysis: the social-educational investigation report (*Recueil de Renseignements Socio-Éducatifs*, RRSE) attached to each case is recorded as a structured object with two ordered placement-recommendation fields (the educator's narrative recommendation and the formal post-investigation decision after hierarchical validation within the educator service); a more detailed set of demographic and family attributes of the minor is observed (status as an unaccompanied foreign minor, status as homeless, country of birth, number of siblings, family configuration); and the linkages between measures, judicial events and the institutional gateway (*déférement*, prepared accueil) are stored explicitly. The trade-off is squarely on the time dimension. The report is attached to the measure it concerns only from 2021, when the corresponding module of the information system was deployed: the linkage rate is zero until 2020 and stable at 34 to 41% of placements thereafter. Appendix B describes that subset. Linkage concentrates on the cases the report is written for, older minors with heavier trajectories, and the department of the court and the year of the measure explain more than twice as much of it as the characteristics of the minor (Table B6). Among placements the two groups are close on age, sex and prior custodial exposure, and differ mainly on the length of the judicial history. Everything that uses the recommendation therefore rests on measures indexed between 2021 and 2024, and the long-horizon outcomes are mechanically truncated by the data end and by the eighteenth-birthday censoring.

Sample selection. On GAME and IMAGES the analytical sample is restricted to spells indexed between 2012 and 2022, an eleven-year window that covers both the measure and the detention side. The lower bound is dictated by the detention sub-base: before 2012, detention episodes are recorded sparsely enough that any outcome involving detention is mechanically attenuated. The upper bound is the end of the data. On PARCOURS the sample runs from 2021, the first year in which the social-educational report is linked to the measure it concerns, to 2024, again the end of the data.

Within the GAME and IMAGES window, three structural exclusions restrict the spell base: civil-law procedures, which correspond to child-protection rather than penal proceedings; investigation measures, which are diagnostic exercises prescribed alongside other measures rather than as alternatives to a placement; and administrative wrappers, the supervision-style spells that double-count a placement when the legal vehicle and the substantive intervention generate two rows. B.3 states how rows are collapsed into spells on each side and gives the exclusion rules. The choices tied to the empirical design, described in Section 4, further reduce the structural base.

The unit of observation is the spell episode constructed above. On GAME and IMAGES the full spell base contains 1,802,188 episodes indexed between 2012 and 2022; applying the three structural exclusions leaves 934,374 spells, which form the structural sample on which the empirical strategy operates. On PARCOURS the base contains 980,197 episodes, of which 690,233 are not investigation measures; 80,201 of these carry a linked social-educational report with at least one recorded recommendation, and 4,302 of them a recommendation in favour of

a closed educational centre, which delimits the population on which the recommendation-based comparison is run.

3.2 Variables of Interest

Treatment. The treatment is a minor's first ever placement in a closed educational centre, indexed inside the analytical window of the source used. I limit the treatment to the first placement in a closed centre, because any event that follows it is itself endogenous to that first placement. On the registers, the comparison pool is subject to the same rule: any spell that follows a prior closed-centre placement is excluded from both arms, so no comparison spell belongs to a minor already exposed to the measure at the time of the indexed spell.

Educator's recommendation. On PARCOURS I observe the placement the educator advises the judge in two ways for each measure whose attached social-educational report is linked: (i) the narrative recommendation written by the educator in charge, and (ii) the formal decision retained at the end of the report after hierarchical validation within the service. Each is coded as a set of indicators over the response families, so a report may advise more than one: in four cases out of ten the report attached to a placement effectively executed in a closed centre also advises the supervision order that would carry it. This variable enters the analysis in two ways: as a control, and as the definition of the population on which the effect is estimated, *i.e.* the minors the institution judged suitable for the measure.
The 22 pre-spell covariates entering the propensity score and what each outcome horizon imposes on the estimation sample, are defined in B.3.

Outcomes measured from the decision. I derive binary outcomes from the spell base by looking forward from the *decision date* of the indexed spell over a window of fixed length[4]. Within an observable window, the outcome equals one if at least one event of the relevant type occurs. The horizons are 12 and 18 months on both sources. Because the window has the same length on both arms, placed and comparison minors are observed over the same amount of time.

The censoring at age 18 shapes how the horizons should be read. The administrative system that produces the data loses jurisdiction at the minor's eighteenth birthday and stops recording events on the individual, so a fixed window is defined only when it closes before that birthday and before the end of the data. The longer horizons therefore retain the minors placed earliest and drop those placed within a year or two of majority, which is why sample sizes fall from one horizon to the next in every table. The estimates at the longer horizon speak about a younger placed population than the estimates at twelve months, and the two are reported side by side for that reason rather than as a single trajectory over time.

[4]See Appendix F and Section 4.4 for a discussion on why the decision date is preferred as the reference rather than the end of the indexed measure.

3.3 Descriptive Statistics

Table B5 in Appendix B presents the demographic, judicial-history and current-spell characteristics of three cohorts of minors observed in the 2012-2022 window: minors at their first observed placement of any kind (44,082 minors), minors at their first observed closed educational centre placement (13,921 minors), and minors at their first observed detention episode (27,732 minors). The three columns overlap by design (e.g. a minor whose first placement is in a closed educational centre appears in both the first-placement and the first-CEF columns) but they characterise three institutionally distinct routes into the residential and custodial system.

Demographics and judicial history. The three cohorts are overwhelmingly male, from 92.0 to 94.8%. Foreign-born minors account for 20.0% of first placements, 16.7% of first closed educational centre placements and 28.9% of first detentions, a comparison to be read with caution since the coverage of country of birth is non-random and more complete in the detention subbase. The average age at the indexed event is 16.3, 16.2 and 16.8 years, and the average age at the very first observed judicial measure 15.4, 15.0 and 15.5, so a minor reaching a closed educational centre starts being followed about a year and three months before the indexed spell. The histories then diverge sharply. The first-placement cohort carries on average 4.0 prior measures of any kind and 0.5 prior criminal court referrals, against 7.1 and 1.0 for both the closed-centre and the detention cohorts, which look strikingly similar to one another on these aggregate counts.

A shared custodial track record. 45.4% of first-CEF minors and 38.3% of first-detention minors have already had at least one placement of any kind before the indexed spell, and 14.6% against 9.5% a prior reinforced educational centre. Most strikingly, 20.4% of first-CEF minors had a prior detention episode and 17.8% of first-detention minors a prior closed educational centre placement: one in five closed-centre minors arrives with a custodial track record, and one in six detained minors with a closed-centre record. The first-placement cohort is far lighter, with 10.7% prior detentions.

Duration and legal context. A first closed educational centre placement lasts on average 137 days, or 4.5 months, against 139 days for a first placement of any kind and 94 for a first detention, and its duration is far more concentrated than either (standard deviation of 88 days against 139 and 140). The 4.5-month average matches the 4.3 months reported by Direction de la Protection Judiciaire de la Jeunesse (2021) and the distribution documented on the 2003-2007 cohort, roughly one third of minors staying less than four months, one third between four and six, and one third beyond (Bruyn et al., 2012); Section 6 returns to what that dispersion does to the estimates. The legal vehicle in force separates the cohorts more clearly than any other dimension: 45.0% of first-CEF minors and 41.5% of first-detention minors have an active judicial supervision order at the start of the indexed spell, against 22.9% for first placements, with the same pattern on active suspended sentences (12.7 and 17.5 against 5.7%). The centre is therefore prescribed within an ongoing judicial supervision relationship and operates as the

residential reinforcement of an existing coercive vehicle.

Nature of the case. Unlike GAME/IMAGES, PARCOURS records the nature of the offence on the subset of measures whose case is documented, and Table 2.2 reports the offence composition of each response family, weighting each measure by the inverse of the number of measures its minor contributes, so that every minor carries a total weight of one and the same underlying facts are not counted several times when a minor received measures in several families.

The gradient runs in the direction the institutional description would predict: homicide, attempts included, concerns 1.1% of the minors under community supervision, 3.1% of those in another residential placement, 6.1% of those in a closed educational centre and 7.4% of those in detention, and robbery and aggravated theft follow the same ordering, from 24.1 to 37.3%. The centre and the prison nonetheless separate on other families, public order and justice offences being close to twice as frequent in detention, 16.8 against 8.8%, and drug offences more frequent in the centre, 27.2 against 24.8%, which is a first indication that the two responses do not recruit on the same dimension.[5]

Table 2.2: Offence shares by response family, each minor weighted equally

	Community supervision	Residential placement	Closed educational centre	Detention
Sexual offences	0.066	0.071	0.078	0.067
Homicide (incl. attempted)	0.011	0.031	0.061	0.074
Robbery and aggravated theft	0.241	0.263	0.326	0.373
Physical assault	0.233	0.209	0.169	0.172
Threats and harassment	0.040	0.062	0.025	0.030
Theft, handling and fraud	0.085	0.084	0.075	0.069
Criminal damage and arson	0.073	0.057	0.030	0.054
Drug offences	0.221	0.244	0.272	0.248
Weapons offences	0.037	0.018	0.042	0.036
Public order and justice	0.117	0.108	0.088	0.168
Road traffic offences	0.071	0.026	0.018	0.052
Other offences	0.025	0.016	0.005	0.018
Measures	42,939	1,540	454	1,755
Minors (total weight)	18,530	524	132	606

Notes: This table reports the share of measures whose attached case carries at least one offence of the indicated family, by the response family the measure belongs to. A minor who received measures in several response families would otherwise contribute the same underlying facts to several columns, which compresses the differences between families. Each measure is therefore weighted by the inverse of the number of measures its minor contributes, so that every minor carries a total weight of one, spread across the families in which they appear. The last two rows report the number of measures each column rests on and the corresponding total weight, that is the equivalent number of minors. The offence nature is recorded for a subset of measures only, and its coverage varies across response families, so the shares describe the documented cases rather than the full flow. Source: SSER data [PARCOURS, 2021-2024].

[5]The offence nature is recorded for 31.1% of closed educational centre measures, 38.0% of other residential placements, 63.2% of community supervision measures and 88.9% of detentions, so coverage is uneven across responses. It is not selective on the minors: measures with and without a recorded offence differ by less than 0.08 standardised deviations on age, sex, prior judicial episodes, prior placements, prior detentions and the two recommendation indicators. Coverage also rises from 48.3% of measures in 2021 to 76.0% in 2024, consistent with the progressive deployment of the recording module rather than with case selection.

Three observations matter for the empirical strategy. First, minors who reach a closed educational centre on average do not look like the broader population of placed minors; they look closer to detained minors in volume of priors, intensity of concurrent supervision and the share of prior detentions, and only slightly younger and less foreign-born. Second, the closed-centre and the detention cohorts share a sizeable revolving door: roughly one in five minors of either cohort has already received the other measure type. Third, the closed-centre placement is rarely a first contact with the system; it is overwhelmingly a renewal or an escalation within a longer judicial trajectory. Identifying the causal effect of the closed-centre placement therefore requires a counterfactual that resembles the treated minors on this trajectory rather than on the full distribution of placed minors.

4 Empirical Strategy

4.1 Identification Challenge

The central empirical challenge is that minors are not randomly assigned to a closed educational centre. Judges select treated minors on a wide set of dimensions: age at decision, prior judicial history and prescriber department are observed in the main administrative source, but the substantive nature and severity of the offence, the minor's family situation, the educator's placement recommendation to the judge, the prosecutor's pressure for a custodial response and the judge's own risk tolerance are not available for the period 2012-2022[6]. A naive comparison of CEF minors with all other placed minors would therefore confound the effect of the placement itself with the selection process.

Four designs the literature uses elsewhere are not available here, for reasons related to the nature of the data. A staggered difference-in-differences design on the geographical roll-out would need the early years of that roll-out to be observed, and they are not: the prescriber department is partly missing in 2004-2006, the volume of recorded measures stabilises only around 2007, and the detention sub-base is almost empty before 2010 and complete only from 2012 (Appendix B). Pre-trends estimated on that baseline are not credible, and detention outcomes are mechanically attenuated for the cohorts treated early.[7] A judge design cannot be built because the magistrate's identifier is not recorded, and a queue or capacity design has nothing to work with either, since closed centres run at about 67% of their authorised capacity nationally (Cour des comptes, 2023) and no saturation displaces minors across facilities (Section 2.3 returns to the part played by the availability of places).

Instrumenting placement by the local availability of a centre in the prescribing department

[6]The educator's recommendation and the nature of the offence are recorded only in the PARCOURS dataset, on a subsample for the post-2020 sub-sample and are exploited for this period and to check the main estimates against a richer conditioning set.

[7]An effort is currently underway to recover individual-level detention records from the *"Direction de l'Administration Pénitentiaire"* (DAP), covering both the period before 2012 and the post-majority years that the DPJJ system stops recording at the minor's eighteenth birthday. The DAP records would close both coverage gaps and, in particular, allow the staggered design to be revisited on a longer and more complete panel and the outcomes to be tracked into early adulthood. The empirical strategy below relies on the existing DPJJ data as of the analytical window 2012-2022.

is not possible because among placements ordered by a court whose department operates a centre, only 34.5% are executed there[8], which is hard to reconcile with a distance-cost channel and consistent with a salience one. Comparing placed minors with detained minors only would carry a problem of its own, since one in five minors entering a closed centre for the first time has already been detained and one in six minors entering detention for the first time has already been in a closed centre, so taking detention as the sole counterfactual would fold the displacement from detention together with every other displacement the placement operates.

The strategy adopted in this paper therefore proceeds in two steps. The first step discusses the nature and construction of the counterfactual, which is inherently dependent on the way the decision to place the minor in the closed educational centre is made by the judicial system. For all the reasons listed above, this construction is not obvious. I explore the decision-process of the placement decision by documenting who is sent there, by whom, and whether that decision follows a recognisable rule using both the PARCOURS database which covers the 2021-2024 period with detailed information such as the educator's placement recommendation to the judge and, in some cases, the nature of the offence, as well as the GAME/IMAGES database covering the 2012-2022 period, less documented but allowing more sample power. This allows me to work on a credible counterfactual group of similar minors. The second step documents the effects of a first CEF placement on subsequent judicial outcomes using different designs, which differ in the information they exploit. Using PARCOURS, the comparison group is made of the minors the institution judged suitable for the measure and the judge redirected elsewhere, and the recommendation and the recorded nature of the offence enter as controls, so that selection on what the educator observed can be tested directly. On GAME/IMAGES, where neither is recorded, the comparison group is constructed from the pre-spell trajectory: I restrict the sample to the region of common support and reweight the untreated spells so that their covariate distribution matches that of the placed minors, first pooling all alternatives and then separately against each type of measure the placement displaces. Section 4.2 defines the counterfactual, Section 4.3 sets out how I document the allocation to the closed educational centre, and Section 4.4 describes the estimation.

4.2 Defining the counterfactual

Importance of the decision process. The effect of a closed educational centre placement can only be read against the measure it replaces, and the law does not fully settle what that measure is. The 2002 legal text presents the closed educational centre as an alternative to incarceration, but whether it operates that way is an empirical matter that determines both the estimand and the interpretation of the coefficients. If we think that the severity of punishment is a positive linear function of the intensity of the judicial history of the child as Figure A2 presents, the centre stands immediately below prison and judges route towards it the minors they would otherwise detain, and the relevant comparison is with detention. If instead the centre is a residential intensification of an ongoing supervision situation, it absorbs minors who would

[8]This number rises to more than 60% at the regional level but there is no available variation after 2010 because every region has been treated.

otherwise have had community-based educational measures, and the relevant comparison is with community supervision. The two views imply different counterfactuals and different policy conclusions, so the allocation is characterised before any effect is estimated. This characterisation is also a result in itself: If the allocation is only weakly predicted by the trajectory the courts record, the minors who reach the centre are not necessarily those for whom the policy was designed. The intervention may be harmless for the population it was built for and harmful for the population it receives.

Characterisation of the decision process. Let D_i denote the response type received by minor i at the indexed spell,

$$D_i \in \{\text{community supervision},\ \text{residential placement},\ \text{closed centre},\ \text{detention}\},$$

and let $\boldsymbol{X}_i$ be the vector of pre-spell characteristics described in Section 3 - demographics, age at first judicial contact, cumulative counts of prior measures by type, prior detentions, and the coercive vehicle in force at the decision. I estimate the following multinomial logit:

$$\Pr(D_i = k \mid \boldsymbol{X}_i, \lambda_t) = \frac{\exp(\boldsymbol{X}_i'\boldsymbol{\beta}_k + \lambda_{t,k})}{\sum_m \exp(\boldsymbol{X}_i'\boldsymbol{\beta}_m + \lambda_{t,m})}, \qquad k \in \{1, \ldots, 4\}, \tag{2.1}$$

with community supervision as the reference alternative and λ_t a set of year effects. The coefficients $\boldsymbol{\beta}_k$, reported as relative risk ratios, say on which dimensions the system sorts. Two patterns are possible: (i) if a characteristic raises the relative risk of both the centre and detention with comparable magnitudes, the two responses are ordered on a single severity dimension and the detention counterfactual is the right benchmark; (ii) if a characteristic raises one and lowers the other, they are separate tracks. I also estimate an ordered logit imposing a single latent severity index and compare the fit, since the ordered model is the formal statement of the one-dimensional view. Section 4.3 defines the two statistics through which I read the predictive content of equation (2.1).

Characterisation of the counterfactual. The judge decides on the observable trajectory and on unobserved information that the administrative sources do not easily record. The PARCOURS database gives me access to that second component through the social-educational report described in Section 3, which states the placement the educator advises and, on a subsample, the nature of the offence. The recommendation summarises what the educator saw and the register omits, so conditioning on it removes part of the selection no trajectory variable can capture; Section 5.1 measures how much.

The counterfactual is therefore drawn from the minors the institution judged suitable for the measure. I take the set of minors for whom the educator recommended a closed educational centre and follow what actually happened to them: some are placed, the remainder are redirected elsewhere. The redirected minors form the comparison group, since the institution judged them suitable for the same measure on the same kind of information and the judge decided otherwise. The estimand is therefore the effect on minors who were both recommended for a closed centre

and placed in one, and not the effect on the placed population as a whole. Restricting the comparison to the recommended population changes which minors the estimate speaks about; it carries no implication for the sign of whatever residual selection remains. I also do the opposite to answer the targeting question: conditional on having been placed, what did the educator recommend? If a majority of placements carry no closed-centre recommendation, the population reaching the centre is not the population the institution identified for it.

Implications of data limitations. In PARCOURS, the report is linked to the measure it concerns only from 2021, so the recommended population is thin and its post-placement horizon is truncated by the end of the data and by the eighteenth birthday. GAME/IMAGES records cover 2012-2022 with an order of magnitude more placements and a genuine post-placement horizon, at the cost of observing neither the recommendation nor the offence. Once the recommendation is shown to capture a substantial part of what the administrative trajectory misses, an estimate built without it cannot be presumed unbiased: the recommended population is where the comparison can be defended, and the estimates obtained there are the ones the design supports best. GAME/IMAGES provides is the statistical power and the horizon that population lacks. I therefore read its estimates as a body of converging evidence rather than as a causal estimate in their own right. Therefore, for the full period the comparison group is constructed rather than derived from the institution, using the following propensity score:

$$p(\boldsymbol{X}_i) = \Pr(CEF_i = 1 \mid \boldsymbol{X}_i, \lambda_t), \tag{2.2}$$

where CEF_i indicates a first closed-centre placement, by logit on the same covariates and year effects, and restrict the sample to spells with $p(\boldsymbol{X}_i) \geq p_{10}$, the tenth percentile of the score among placed minors. The threshold is what defines the common support here: it is computed on the placed only, and then applied to placed and untreated spells alike, so both arms are restricted to the same region of the score. The rule discards the tenth of placements whose observable profile is least distinguishable from untreated spells, together with untreated spells below that threshold; it retains 90% of placements and defines the population on which every pooled estimate is computed. Figure C1 reports the score distribution of the two arms and the position of that threshold, Table C1 the covariate means before and after reweighting, and Figure C2 the standardised differences the reweighting achieves. Section 5 reports what the untreated spells of that population actually receive, which is the GAME counterpart of the destinations observed in PARCOURS, so that the two counterfactuals can be compared directly. I discuss in more details the limitations of this approach, and the ways I check for coherence, sensitivity and robustness.

4.3 Documenting the allocation to a closed educational centre

The unexplained part of the decision. Characterising the counterfactual establishes which measures the placement displaces, and leaves open how much of the decision the recorded trajectory accounts for. However, that share matters for identification. What the trajectory does

not explain is either information the judge holds and the data do not record, in which case placed and unplaced minors differ on grounds I am not able to account for, or institutional noise unrelated to the minor, in which case the residual is harmless. I quantify the predictable part with two statistics computed from equation (2.1): the area under the receiver operating characteristic curve, which measures whether the model orders minors correctly, and the share of actual placements falling in the top decile of predicted probability, which measures whether it identifies them. The two answer different questions, as a decision may be well ordered without being targeted.

The GAME/IMAGES records cannot distinguish hidden information from noise, since both appear as unexplained variation. PARCOURS can, because it records what the institution itself concluded from the case file. I re-estimate the same choice model on the report-linked sample, adding the twelve offence families first and the educator's recommendation second, and read the movement in the two statistics. If predictability is restored once I control for the recommendation, the residual of the administrative model was information the institution possessed, and the estimation must condition on it, which is what the PARCOURS design does. If it moves little, the residual is not about the minor at all, and attention shifts to the process by which a minor is matched to a place.

Logistical evidence. The second reading can be tested on its own, through evidence of the matching process to a closed centre rather than of the case. First, I rank, for each placement, all centres open that year by their distance to the prescribing court, and I record the rank of the one that admitted the minor. The frequent use of more distant centres[9] shows that proximity does not fully determine allocation, despite it being an explicit criterion (up until 2016). Then, I compare the occupancy of the nearer centres that were not retained with that of the centre that admitted the minor, in the month of the placement, which separates a capacity constraint from an admission decision. I also compare facilities on the severity of the minors' profiles sent to them by the same courts in the same years, and asks whether those positions persist across years, which distinguishes selective admission from allocation as places come free. Finally, I turn to the group a minor joins rather than the facility, holding the observed calendar of admissions fixed. I reassign entrants at random to those dates, so that similarity between minors entering together can be compared with what timing alone would produce. However, none of these exercises identifies who selects because a facility declining a profile and a placing service anticipating that refusal leave the same trace, and because the solicitations themselves are not recorded.

4.4 Estimating the effects of placement in a closed educational centre

Estimation within the recommended population. On PARCOURS, the estimating sample is the set of measures whose attached report recommended a closed educational centre. For minor i with outcome Y_i, I estimate the following model:

[9] I discuss measurement limitations of the distance more precisely in Section 5.1

$$Y_i = \alpha + \beta CEF_i + \boldsymbol{X}_i'\boldsymbol{\gamma} + \mu_c + \lambda_t + \varepsilon_i, \tag{2.3}$$

where CEF_i indicates effective closed educational centre placement, $\boldsymbol{X}_i$ collects the demographics and prior-history controls, μ_c are fixed effects for the department of the prescription court and λ_t for the year of the measure, with standard errors clustered at the court-department level. Without the set of covariates $\boldsymbol{X}_i$, β is the raw contrast between placed and redirected minors, inside the population of minors who have been recommended for a CEF placement. Adding $\boldsymbol{X}_i$ asks whether the difference survives the observable judicial trajectory of the minor. I then add the two recommendation indicators to $\boldsymbol{X}_i$: if placement was driven by what the educator saw and I do not, β should fall substantially once the recommendation is held constant, so a stable coefficient is evidence that the recommendation does not add substantial information to the decision that the detailed set of covariates on the judicial history already does. On the subsample where the nature of the case is recorded, I add twelve offence family indicators, first as controls in equation (2.3) and then in the balancing set of the weights described below, so that placed and comparison minors are matched on what the minor actually did.

Estimation on the constructed comparison group. On GAME/IMAGES, placed and untreated spells differ on the trajectory by construction, so comparability must be imposed. I use entropy balancing (Hainmueller, 2012) so that each untreated spell j receives a weight w_j ensuring that the weighted control mean of every matching covariate exactly equals the unweighted treated mean on the sample of minors placed in closed educational centres. Placed spells keep unit weight, so the estimand is the average treatment effect on the treated. Balance is imposed exactly rather than achieved in expectation, at the cost of a reduction in the effective size of the comparison group, which I report. I then estimate the effects of placement in a closed educational centre by estimating the following model:

$$Y_i = \alpha + \beta CEF_i + \boldsymbol{X}_i'\boldsymbol{\gamma} + \lambda_t + \rho_d + u_i, \qquad \text{weighted by } w_i, \tag{2.4}$$

with λ_t year effects, ρ_d prescriber-department fixed effects and standard errors clustered at the prescriber-department level. Including $\boldsymbol{X}_i$ alongside the weights makes the estimator doubly robust, meaning that β is consistent if either the weighting or the outcome regression is correctly specified.

Estimation against each alternative. Because the counterfactual is composite, a pooled β averages over different types of displacements. I therefore hold the treated group fixed and vary the comparison pool: for each type of alternative measure in turn, the pool is restricted to spells of that type and entropy balancing is applied so that the restricted pool matches the same treated profile. Each panel answers what would have happened for the population of placed minors had they received that alternative, and documents how β varies with what the placement displaces.

Sensitivity to the adjustment. To make sure that the estimates are not an artefact of the weighting scheme, I report how they vary on the same sample for multiple specifications: the raw contrast with fixed effects only; the same with controls for $\boldsymbol{X}_i$ and no weights; inverse-probability weights $w_j = p(X_j)/(1-p(X_j))$ for the comparison spells, the classical propensity-score weights for the effect on the treated, with and without $\boldsymbol{X}_i$; and entropy balancing with and without $\boldsymbol{X}_i$. Agreement between the two weighting families indicates that β does not depend on how the propensity score is used. I also vary the support rule of equation (2.2), from no restriction to increasingly demanding overlap, so that sensitivity to the definition of the comparison population is visible. Finally, the support of equation (2.2) is estimated once on the full pool, so a comparison against a particular type of measure inherits a threshold computed against all alternatives at once. I therefore also re-estimate the score and the support separately within each comparison, so that the retained region is the one where placed minors and that specific type of alternative actually overlap; Table E1 reports these estimates.

Timing of the outcome window. Outcomes are measured over fixed windows running from the decision date: an event counts if it occurs strictly after the decision and within 12 or 18 months of it, depending on the horizon. Indeed, the decision is the only date predetermined on both arms, so placed and comparison minors are observed over exactly the same length of time; and every subsequent event counts, including a custody that revokes the coercive vehicle while the placement is still running. Additionally, revocation is one of the ways the placement produces custody, and the legal device that makes it possible exists only on the placed arm. Measuring from the decision therefore estimates what follows from the decision to place, which is the quantity a court would want to know.

The alternative convention, which opens the window at the end of the indexed measure, is reported in Appendix F, and answers what happens once the measure is over. However, the end of a measure is not a neutral date: a supervision order that fails ends on the day the minor is incarcerated, whereas one that runs its course lasts two years, so the terminal date is partly produced by the outcome being measured. This differs systematically across arms, which makes the observation windows unequal. The median closed educational centre placement lasts 159 days against 312 for the eligible controls as a whole and 498 for a standalone supervision order, and the minor is 16.5 years old when the placement closes against 17.1 for the controls, so an outcome accumulated up to the eighteenth birthday is accumulated over roughly half again as much time on the placed arm. Table B4 in Appendix B reports these durations by type of alternative measure. Section 6 returns to how much of the estimate travels through failure during the stay.

5 Results

5.1 The decision process of the placement

Who decides, and on what. The decision to place in a closed educational centre is taken by a judge on a case file. To first understand the decision of placement, I look at which magistrate decides and on what grounds. Table 2.3 reports the distribution of placements by prescribing authority. The juvenile judge orders 72.6% of CEF placements, the liberty and custody judge has the most intensive use of CEF placement at 13.0% of the spells handled, and the prosecutor never places, with zero placements out of 54,333 spells. This implies that the placement is mainly a decision of the bench, taken on a file, and not a pure administrative allocation.

Whether the centre stands immediately below prison in the placement decision process is then tested by the multinomial logit of equation (2.1), whose relative risk ratios are reported in Table D1 of Appendix D.

If the two responses belonged to a single severity ordering, the characteristics raising the relative risk of a placement would also raise that of detention, and several of them do the opposite. Age moves in opposite directions, an additional year lowering the relative risk of a placement to 0.60 and raising that of detention to 1.10.

Table 2.3: Judicial responses by prescribing authority

Prescriber	Spells	Community supervision	Residential placement	CEF	Detention	Share of all CEF placements
Court of appeal	603	40.8%	21.4%	11.6%	26.2%	0.5%
Juvenile judge	317,811	75.9%	16.8%	2.9%	4.4%	72.6%
Investigating judge	46,832	56.8%	14.2%	4.0%	25.0%	14.6%
Liberty and custody judge	6,177	33.7%	29.3%	13.0%	24.1%	6.2%
Other prescriber	1,017	60.9%	2.7%	1.6%	34.9%	0.1%
Public prosecutor	54,333	100.0%	0.0%	0.0%	0.0%	0.0%
Juvenile court	125,156	93.1%	2.7%	0.6%	3.6%	5.9%

Notes: This table reports, for each prescribing authority, the number of spells prescribed in the 2012-2022 working sample and their distribution across the four response families; the last column reports each authority's share of all first closed educational centre placements. The public prosecutor essentially never prescribes a CEF placement, which is institutionally consistent: the placement can only be ordered by a judge, as an accessory to a coercive judicial vehicle. Percentages are computed within prescriber. Source: SSER data [GAME/IMAGES, 2012-2022].

An active judicial supervision order multiplies the relative risk of a placement by 4.18 against 2.93 for detention, and a prior reinforced centre by 2.38 against 1.58, while the count of prior episodes weighs far more heavily on detention (1.75) than on the centre (1.05). Imposing a single direction is costly: the ordered logit reaches a McFadden pseudo R^2 of 0.084 against 0.116 for the unordered model on the same covariates, so about a quarter of the explanatory power is discarded, and the ranking survives an information criterion that charges for the extra parameters the unordered model carries. Forced to resolve conflicting movements into one sign, the ordered model concludes that older minors receive less severe responses, which holds for the closed centre and does not for detention. The centre and the prison therefore recruit on different dimensions: the centre takes younger minors already held under a coercive order, the prison older minors with longer records.

This distinction matters for the choice of counterfactual. If the CEF was simply the step immediately below detention on the severity scale, detention would be the natural alternative. The allocation results give little support to that interpretation as CEF placement and detention respond differently to several observed characteristics, and the unordered model fits the data substantially better. They therefore provide no reason to assume that minors placed in a CEF would otherwise have been detained. The relevant alternative has to be established empirically. Section 4.2 does this by documenting the measures received by comparable minors who were recommended for a CEF but not placed there.

How far the decision can be predicted. Knowing on what the system sorts leaves open how much of the decision the data accounts for, which matters for identification: what the judicial trajectory does not explain is either information the judge holds and the data does not include, in which case the comparison of placed and unplaced minors is compromised, or institutional noise, in which case it is harmless. Table 2.4 settles how much of each the two sources can distinguish, by reporting the same two statistics under three nested sets of information: the recorded trajectory alone, the trajectory plus the twelve offence families, and the trajectory plus the educator's recommendation.

Table 2.4: What predicts the judicial response: recorded trajectory, nature of the offence, and the educator's recommendation

		Area under the ROC curve			Share captured by the top decile		
Response	Spells	(1)	(2)	(3)	(1)	(2)	(3)
Panel A: historical registers, 2012-2022							
Community supervision	441,504	0.744	–	–	11.6%	–	–
Residential placement	65,412	0.719	–	–	30.0%	–	–
Closed educational centre	12,848	0.767	–	–	41.0%	–	–
Detention	32,165	0.793	–	–	40.0%	–	–
Panel B: report-linked sample, 2021-2024							
Community supervision	67,917	0.740	–	0.754	10.8%	–	10.9%
Residential placement	4,048	0.710	–	0.710	27.4%	–	27.8%
Closed educational centre	1,460	0.769	–	0.850	37.8%	–	54.8%
Detention	1,974	0.783	–	0.800	38.8%	–	40.7%
Panel C: sub-sample with a recorded offence, 2021-2024							
Community supervision	42,939	0.756	0.774	0.781	10.7%	10.8%	10.8%
Residential placement	1,540	0.709	0.722	0.722	31.4%	33.6%	33.8%
Closed educational centre	454	0.791	0.814	0.877	42.1%	45.8%	60.4%
Detention	1,755	0.794	0.823	0.835	39.1%	40.5%	43.1%

Notes: This table reports how far each judicial response can be predicted from three nested sets of information, using the fitted probabilities of a multinomial model of the response received. Specification (1) uses the judicial trajectory recorded before the decision together with the demographic covariates and year effects; specification (2) adds the twelve offence families; specification (3) adds the educator's narrative recommendation and the formal post-investigation decision. For each specification the table reports the area under the ROC curve of the fitted probability against actual receipt, where 0.5 indicates no discrimination and 1 perfect discrimination, and the share of actual recipients falling in the top decile of the fitted probability distribution. Panel A uses the historical registers, which record neither the offence nor the recommendation, so only specification (1) can be computed there. Panel B uses the full report-linked sample, where the recommendation is observed but the offence is not, so specification (2) cannot be computed. Panel C restricts to the measures for which the nature of the offence is recorded and computes all three. A dash marks a cell the corresponding source cannot produce. Source: SSER data [GAME/IMAGES, 2012-2022 and PARCOURS, 2021-2024].

On the registers, which record neither the offence nor the recommendation, the area under the curve for the closed centre is 0.767: between a placed and a non-placed minor drawn at random, the model ranks the placed one higher in roughly three quarters of pairs.[10] Ranking is not targeting, however. The 10% of minors the model judges most likely to be placed contain only 41.0% of the placements actually made, so a rule built on the recorded trajectory would miss three placements in five, and a minor who was in fact placed carried an average predicted probability of 0.057, so the model ranks them above almost everyone else and still considers placement unlikely for them. Both follow from how rare the measure is, since only 2.33% of spells end in a closed centre. The recorded trajectory therefore says reliably who is more at risk than whom, without describing a profile that would pick out the minors who are placed.

Separating the two readings of that residual requires information the registers do not contain and the successor system does. Adding the educator's recommendation raises the area under the curve for the closed centre from 0.769 to 0.850 and the top-decile share from 37.8 to 54.8%, so a substantial part of what the trajectory missed was information carried in the file and summarised by the institution, which is precisely the channel the estimation of Section 5.2 conditions on. Nothing comparable happens on the other responses: on community supervision the curve moves from 0.740 to 0.754 and on ordinary residential placement it does not move at all.

The most natural objection is that the recommendation is only a proxy for the gravity of the offence, which the registers do not code. The third panel tests it on the sub-sample where the nature of the case is recorded, adding the offence families first and the recommendation only afterwards. For the closed centre the area under the curve moves from 0.791 to 0.814 and then to 0.877, and the top-decile share from 42.1 to 45.8 and then to 60.4%. What the educator brings is therefore not just a reassessment of the charges against the minor. The comparison with detention makes this more obvious, as the offence families contribute more than the recommendation does, raising the area under the curve from 0.794 to 0.823 against a further 0.012 for the recommendation. Detention is decided on what the minor did whereas the closed centre is decided on an assessment of what the minor needs, and that assessment is the part of the decision the administrative trajectory cannot reproduce.

How a placement is matched to a facility. In theory, guidelines ask the placement service to look for a nearby centre, so that the minor keeps contact with the family and the exit can be prepared. In practice the receiving centre is rarely the nearest one. It takes the minor in 24% of cases, the median placement goes to the third nearest, 17% go past the tenth, and the median minor travels 100 kilometres against 27 when the nearest facility takes them, for a mean distance of 133 kilometres (Table D2 of Appendix D, which also gathers the other allocation tables cited in this paragraph and the two that follow). Three placements in four pass at least one centre that was closer and did not take the minor, which fits what Direction de la Protection Judiciaire de la Jeunesse (2020) reports: the median case involves contacting five establishments, and in one case in six the territorial directorate has to step in to find a place. Distance here is measured

[10] Re-estimated on a random half of the sample and evaluated on the other half, the out-of-sample area under the curve is 0.766, so the figure is not an artefact of overfitting.

from the prescribing court and not from the minor's home, which the data do not record, so it says something about how the match is made rather than about the burden on the family.

Two things could explain this without any selection at the admission stage, and neither holds up. The first is the 2016 circular, which replaced a local search giving priority to proximity by a national procedure. The two periods look the same: the median rank of the receiving centre is three on both sides, the mean distance moves from 134 to 132 kilometres, and the share sent to the nearest centre from 24.9 to 22.8% (Table D3). Two of these differences are statistically significant, but they are computed on a very large sample of placements and are too small to matter. Moreover, the distance moves the opposite way from what a nationalisation would produce. The second is congestion. Comparing, in the month of each placement, how full the nearer centres that were passed over were with how full the centre that admitted the minor was, the centres that were skipped are the less loaded ones, by two to four points depending on the rank (Table D4). The friction is on the admission side, which is also why the national occupancy rate can sit between 60 and 70% while judges report difficulty placing.

What the friction does to the decision. Among minors whose report recommended a closed centre, 13.2% are placed in one: 9.0% where the court's department has no centre, 15.2% where it has one (Table D5). Nine recommendations in ten do not end in the measure the institution advised. Why local presence matters when admission is organised nationally is not something the data settle. The friction also shows up in who each facility receives. Comparing facilities on the minors sent to them by the same courts in the same years, the share of entrants with a prior detention episode runs from 8% in the lightest quarter of facilities to 30.5% in the heaviest, and mean age from 15.1 to 16.7 years, while the share of publicly run establishments moves from 0.38 to 0.15 to 0.42 to 0.38 across the same quarters, so what separates facilities is not the sector they belong to (Table D6). These positions last: a facility in the heaviest third one year is there again the next in 43.6% of cases, against the 33% random allocation would give (Table D7), and a quarter of the spread between facilities is permanent (Figure D1). The same shows on the minor's side: at equal court and year, a minor with a one standard deviation heavier profile is placed 0.53 rank further down the queue and travels 8.4 kilometres more (Table D8).

Therefore, there seems to be selection at the admission stage rather than congestion. However, this does not say who selects: a facility turning down a profile and a placing service that anticipates the refusal leave the same trace, and the solicitations are recorded nowhere.

Who enters together. The same logistics govern the composition of the group a minor joins, which matters for two reasons: peer composition is a plausible channel of the effect, and a design exploiting it would require entry timing to be as good as random within a facility. I test that by holding the observed calendar of admissions fixed and reassigning profiles at random to those dates within each facility-year, a thousand times. Half of admissions occur on the same day as another admission to the same centre, and minors entering within thirty days of each other are more alike in prior episodes, in age and in prior-detention status, and

share a prescribing department more often, 19.7 against 15.6%, than under any of the thousand reshufflings (Table 2.5). Two minors admitted within thirty days of each other differ by 7.37 prior judicial episodes on average, against 7.58 had the same dates been filled from that facility's intake in the same year and 7.94 from all its years, so entrants are more alike than chance would make them and the gap widens as the reshuffling reaches further. Who lives with whom seems to be settled by judicial batches rather than by the timing of individual trajectories. This implies that peer-effects cannot be studied. Indeed, the design that identifies peer effects in juvenile facilities elsewhere holds the establishment fixed and reads peer influence off variation in who happens to overlap within it (Bayer et al., 2009), which requires that overlap to be idiosyncratic, and the permutation test rejects exactly that, therefore no correlation between cohort composition and later outcomes can be read as a causal peer effect here.

Table 2.5: Similarity of co-entrants against a timing-permutation placebo

	Observed	Placebo mean (within year)	p-value (within year)	Placebo mean (across years)	p-value (across years)
Absolute difference in prior judicial episodes	7.369	7.579	0.001	7.939	0.001
Absolute difference in age at entry (years)	0.771	0.866	0.001	0.884	0.001
Share with the same prior-detention status	0.621	0.603	0.001	0.591	0.001
Pairs of admissions within 30 days	69,919				
Permutation replications		1000		1000	

Notes: This table compares the observed similarity of pairs of admissions into the same closed educational centre within thirty days of each other (each admission forms one pair with every other admission of the same facility in its thirty-day window, so the number of pairs exceeds the number of stays) with the similarity obtained under two timing placebos in which admission dates are held fixed and the entrants' profiles are randomly reassigned to them: within each facility-year, which preserves each year's intake composition and breaks only the fine timing, and within each facility across all years, which additionally breaks year-to-year drift in the intake. The first two rows report the mean absolute difference within the pair, the third the share of pairs agreeing on prior-detention status. Each p-value is the one-sided probability, across one thousand reshufflings, of a similarity at least as strong as observed. The sample covers terminated 2012-2022 stays with an identified facility and matched pre-entry characteristics. Source: SSER data [GAME/IMAGES, 2012-2022].

Taken together, these results indicate that the counterfactual is not solely detention, since the centre and the prison recruit on different dimensions, which is why the effect is estimated separately against each type of alternative measure, one comparison at a time, rather than against a single pooled comparison group. Selection operates intensely on the recorded trajectory, which is controlled for, plus a residual that the recommendation captures in large part and that is demonstrably not information about the offence. What remains seems governed by an admission process that sorts on the facility side rather than by the local supply of places, which is both why no geographic instrument works here and why I argue the residual is closer to institutional noise than to a hidden severity ranking.

5.2 The effects of the placement

Composition of the two comparison groups. Before turning to magnitudes, I describe the two counterfactuals since the estimates are only interpretable against them. Table 2.6 reports where minors recommended for a closed educational centre and not placed in one actually went: community supervision receives the large majority of them, residential placement a small share, and detention under a tenth.

Table C2 reports the corresponding distribution on GAME, that is, the measures actually received by the untreated spells of the common-support sample of the propensity score. The two distributions agree in shape, which matters: the comparison group constructed statistically on the long sample resembles the one the institution itself designates on the short one. In both cases the placement displaces community supervision far more often than detention, so the policy does not operate as the alternative to incarceration its statute describes.

Table 2.6: Destination of minors recommended for a CEF but not placed in one

	Educator recommendation		Formal post-investigation decision	
Destination	Measures	Share	Measures	Share
Community supervision				
Judicial supervision only	789	21.1%	505	37.4%
Pre-trial supervised release	30	0.8%	4	0.3%
Suspended sentence (SME)	366	9.8%	140	10.4%
Supervised release (LS)	814	21.8%	245	18.1%
Article 16bis measure	0	0.0%	0	0.0%
Sanctions and adjustments	240	6.4%	110	8.1%
Other community supervision	882	23.6%	205	15.2%
Residential placement				
Residential placement	220	5.9%	91	6.7%
Reinforced centre (CER)	39	1.0%	11	0.8%
Detention				
Detention	353	9.5%	39	2.9%

Notes: This table reports, among penal measures whose attached social-educational report recommended a closed educational centre placement but whose minor was not placed in one, the distribution of the measure effectively received. Destinations are aggregated into the same measure families as the composition table of the constructed comparison group, so that the two counterfactuals can be read side by side; the underlying nomenclature is finer and is reported separately. Because the 2021 code reorganised the nomenclature, two correspondences are approximate: the educational judicial measure and its provisional form are mapped onto supervised release, which they replaced, and investigation measures are counted as other community supervision. The first column block conditions on the educator's narrative recommendation; the second on the formal decision retained at the end of the report after hierarchical validation. This non-placed population constitutes the institutionally defined counterfactual group used in the estimation below. The sample covers PARCOURS penal measures with a linked report containing at least one recorded recommendation. Source: SSER data [PARCOURS, 2021-2024].

Table 2.7 runs the mapping in reverse and asks, among effective placements, what the report had recommended. A minority carry an explicit closed-centre recommendation; judicial supervision and an unspecified placement dominate the recommendations attached to minors who nonetheless end up in the centre. The population reaching the facility is therefore not the population the institution identified for it, which is the targeting result the effect estimates below will qualify.

Effects within the population the institution deemed suitable. The most appropriate comparison available is among minors whom the educator recommended for a closed centre: I compare those who were placed with those whom the judge sent elsewhere, using equation (2.3) with outcomes measured over the fixed 12- and 18-months windows starting at the decision date described in Section 4.4. This comparison asks what follows from the decision to give a CEF

Table 2.7: Measures recommended in the report attached to minors placed in a closed educational centre

	Educator recommendation		Formal post-investigation decision	
Measure recommended in the attached report	Placements	Share	Placements	Share
Placement				
Closed educational centre (CEF)	569	39.0%	393	26.9%
Placement, category unspecified	245	16.8%	124	8.5%
Collective housing unit	45	3.1%	20	1.4%
Reinforced educational centre (CER)	18	1.2%	8	0.5%
Child welfare placement	6	0.4%	2	0.1%
Foster family	3	0.2%	0	0.0%
Diversified housing unit	0	0.0%	0	0.0%
Community-based measures				
Judicial supervision	616	42.2%	510	34.9%
Educational judicial measure	170	11.6%	86	5.9%
Detention				
Detention	0	0.0%	124	8.5%
No recommendation of these types				
None of the above	280	19.2%	618	42.3%
Juvenile detention facility (EPM)	0	0.0%	31	2.1%
Remand prison	0	0.0%	17	1.2%

Notes: This table reports, among PARCOURS penal measures effectively executed in a closed educational centre between 2019 and 2024, the measures that the attached social-educational report had recommended. The report contains two ordered fields: the recommendation written by the educator in the narrative section, and the option formally retained at the end of the report after hierarchical validation within the educator service; each is reported in a separate pair of columns. Shares are computed over all placements with a linked report and sum to more than one hundred per cent when a report mentions several options. By construction every measure in the sample carries at least one recorded narrative recommendation, so the shares describe which measure the report advocated rather than whether it advocated one; a placement whose report recommended judicial supervision is a placement the institution had identified for a community-based measure. The final row counts reports whose recommendations fall outside the categories listed. The sample covers measures whose report could be linked, a linkage recorded from 2021 onwards. Source: SSER data [PARCOURS, 2021-2024].

placement among minors the institution itself considered suitable for one (Table 2.8).[11]

Within this population, CEF placement raises the probability of detention within twelve months by 8 percentage points, relative to a counterfactual mean of 43.5% (Panel A). At eighteen months, the estimate falls to 4.2 points and is not statistically significant. The 18-month sample is roughly half as large, so the data does not allow me to distinguish attenuation over time from lower precision. Using the formal post-investigation recommendation instead gives a similar twelve-month estimate of 9.7 points (Panel B), although the smaller sample makes the remaining coefficients less precise. Restricting the comparison group further to recommended minors who received judicial supervision without a CEF yields a smaller positive estimate of 5.5 points at twelve months, which is not statistically significant (Panel C).

The other outcomes help qualify the detention result. Placement reduces the probability of another CEF placement over both horizons, although part of this difference follows mechanically from treated minors spending the beginning of the follow-up period inside a CEF and therefore not being at risk of a new placement. I find no meaningful increase in new court referrals: the

[11] The equivalent specification anchoring outcomes at the end of the measure rather than at the date of the decision, and therefore excluding custody that revokes an ongoing placement, is reported in Table F1 of Appendix F.

12-months estimate is 1.1 points and the 18-months estimate is −2.4 points on control means of 94.9 and 97.2%. The increase in detention therefore occurs without a corresponding increase in the extensive margin of subsequent court referral.

Table 2.8: Effect of a CEF placement within the recommended population, outcomes anchored at the decision (PARCOURS, 2021-2024)

	Detention episode		New CEF placement		New court referral	
	12 months	18 months	12 months	18 months	12 months	18 months
Panel A: Educator recommendation sample						
Placed in a closed educational centre	0.080***	0.042	-0.056**	-0.060*	0.011	-0.024*
	(0.027)	(0.034)	(0.028)	(0.033)	(0.016)	(0.013)
Counterfactual mean	0.435	0.567	0.258	0.331	0.949	0.972
Observations	1,737	897	1,737	897	1,737	897
Treated observations	286	163	286	163	286	163
Adjusted R^2	0.230	0.325	0.198	0.274	0.120	0.214
Panel B: Formal decision sample						
Placed in a closed educational centre	0.097**	0.015	-0.022	-0.082	-0.003	-0.011
	(0.042)	(0.036)	(0.039)	(0.057)	(0.021)	(0.026)
Counterfactual mean	0.401	0.559	0.236	0.323	0.943	0.941
Observations	784	412	784	412	784	412
Treated observations	196	110	196	110	196	110
Adjusted R^2	0.327	0.443	0.293	0.360	0.364	0.571
Panel C: Judicial supervision as the comparison						
Placed in a closed educational centre	0.055	0.017	-0.080***	-0.107**	-0.006	-0.021
	(0.035)	(0.033)	(0.027)	(0.042)	(0.015)	(0.015)
Counterfactual mean	0.470	0.568	0.291	0.385	0.967	0.973
Observations	660	371	660	371	660	371
Treated observations	286	163	286	163	286	163
Adjusted R^2	0.213	0.308	0.153	0.187	0.120	0.221
Included in all regressions:						
Pre-spell judicial history and demographic controls	Y	Y	Y	Y	Y	Y
Year fixed effects	Y	Y	Y	Y	Y	Y
Department fixed effects	Y	Y	Y	Y	Y	Y

Notes: This table reports weighted least squares estimates of the average effect of a first closed educational centre placement on the treated. Each panel restricts the control group to the indicated type of alternative measure; unless stated otherwise the treated group is identical across panels and the entropy-balancing weights are re-estimated within each panel so that the restricted control pool matches the treated covariate profile (weighting method by panel: Educator recommendation sample: Entropy; Formal decision sample: Entropy; Judicial supervision as the comparison: Entropy). Each column corresponds to one post-spell outcome and horizon; outcomes at twelve months are defined only for spells whose full one-year post-spell window is observable before the data end and the minor's eighteenth birthday, and therefore over-represent minors placed young. The counterfactual mean is the weighted control mean of the outcome. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, in parentheses, are clustered at the prescriber-department level. The sample is restricted to penal measures whose attached social-educational report recommended a closed educational centre placement (Panel A: educator narrative recommendation; Panel B: formal post-investigation decision). Treated units are measures effectively executed in a closed centre; controls are recommended measures redirected to any other response. Panel C narrows the control group of Panel A to recommended minors who received a judicial supervision order in place of the placement. Outcomes are measured over fixed windows of twelve and eighteen months running from the date of the decision, so that both arms are observed over the same length of time and every subsequent event counts, including a custody that ends the measure while it is still running. A window is defined only where it is fully observable before the end of the data and the minor's eighteenth birthday, which is why the samples differ from those of the end-of-measure tables. Detention episodes merge consecutive custodial records at most three days apart. These estimates incorporate the revocation channel and answer a different question from the end-of-measure estimates reported in the appendix. Controls include the seventeen demographic, family and prior-history covariates of the PARCOURS analysis; the department is that of the competent juvenile court. Source: SSER data [PARCOURS, 2021-2024].

The within-recommended estimate is smaller than the estimate obtained on the full report-linked sample. Table 2.9 first shows how much of this difference can be related to information contained in the educator's assessment. On the full sample, adding the recommendation indicators reduces the twelve-month detention coefficient from 17 to 14 percentage points. The recommendation therefore captures relevant selection that is absent from the administrative

trajectory, but accounting for it leaves a large positive estimate.

Table 2.9: Placement effect on the full report-linked sample, with and without recommendation controls (PARCOURS, 2021-2024)

	Detention episode		New CEF placement		New court referral	
	12 months	18 months	12 months	18 months	12 months	18 months
Panel A: Covariate adjustment only						
Placed in a closed educational centre	0.170***	0.136***	0.049***	0.055***	0.055***	0.031***
	(0.018)	(0.021)	(0.015)	(0.020)	(0.009)	(0.009)
Counterfactual mean	0.339	0.407	0.201	0.262	0.870	0.887
Observations	29,496	15,283	29,496	15,283	29,496	15,283
Treated observations	717	405	717	405	717	405
Adjusted R^2	0.288	0.364	0.157	0.210	0.137	0.146
Panel B: Adding the recommendation indicators						
Placed in a closed educational centre	0.140***	0.094***	0.052***	0.060**	0.047***	0.025***
	(0.018)	(0.025)	(0.018)	(0.027)	(0.008)	(0.009)
Counterfactual mean	0.339	0.407	0.201	0.262	0.870	0.887
Observations	29,496	15,283	29,496	15,283	29,496	15,283
Treated observations	717	405	717	405	717	405
Adjusted R^2	0.291	0.369	0.157	0.211	0.138	0.146
Included in all regressions:						
Pre-spell judicial history and demographic controls	Y	Y	Y	Y	Y	Y
Year fixed effects	Y	Y	Y	Y	Y	Y
Department fixed effects	Y	Y	Y	Y	Y	Y

Notes: This table reports weighted least squares estimates of the average effect of a first closed educational centre placement on the treated. Each panel restricts the control group to the indicated type of alternative measure; unless stated otherwise the treated group is identical across panels and the entropy-balancing weights are re-estimated within each panel so that the restricted control pool matches the treated covariate profile (weighting method by panel: Covariate adjustment only: Entropy; Adding the recommendation indicators: Entropy). Each column corresponds to one post-spell outcome and horizon; outcomes at twelve months are defined only for spells whose full one-year post-spell window is observable before the data end and the minor's eighteenth birthday, and therefore over-represent minors placed young. The counterfactual mean is the weighted control mean of the outcome. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, in parentheses, are clustered at the prescriber-department level. The sample covers all PARCOURS penal measures with a linked social-educational report containing at least one recorded recommendation; treated units are measures effectively executed in a closed educational centre and controls are all other measures. Both panels use identical entropy-balancing weights computed on the seventeen covariates; the second panel adds, as regression controls, the educator's narrative recommendation and the formal post-investigation decision in favour of a closed centre, both recorded before the judge's decision. A stable coefficient across panels indicates that selection on educator-observed unobservables does not drive the estimate. The department is that of the competent juvenile court. Source: SSER data [PARCOURS, 2021-2024].

The remaining difference between the 8-point within-recommended estimate and the larger full-sample estimate can have two interpretations. The tighter comparison may remove selection that remains in the full sample, or the effect of placement may genuinely be smaller among minors for whom a CEF was recommended. I examine the latter possibility by interacting placement with recommendation status in Table 2.10. At twelve months, the interaction with the educator's narrative recommendation is -5.5 points with a standard error of 4.1, and is therefore too imprecise to establish differential effects; at eighteen months it is -6.1 points and significant only at the 10% level. The estimates are compatible with a smaller effect among recommended minors, but do not distinguish treatment-effect heterogeneity from the additional selection removed by restricting the comparison to that population. I therefore retain the within-recommended estimate as a separate, more strictly conditioned comparison rather than interpreting the difference across samples as treatment-effect heterogeneity.

The main remaining observed dimension of selection is the offence itself. Table 2.11 restricts the sample to cases for which offence information is recorded and adds the twelve offence families first as controls and then to the balancing set. Once the recommendation is included, the 12-

Table 2.10: Heterogeneity of the placement effect by recommendation status (PARCOURS, 2021-2024)

	Detention episode		New CEF placement		New court referral	
	12 months	18 months	12 months	18 months	12 months	18 months
Placed in a closed educational centre	0.158***	0.098***	0.045*	0.044	0.044***	0.026**
	(0.025)	(0.030)	(0.023)	(0.030)	(0.011)	(0.013)
Placed × recommended (narrative)	-0.055	-0.061*	-0.063	-0.052	0.006	0.001
	(0.041)	(0.036)	(0.045)	(0.067)	(0.019)	(0.019)
Placed × recommended (formal decision)	0.009	0.077	0.122***	0.143***	0.002	-0.003
	(0.055)	(0.050)	(0.032)	(0.054)	(0.023)	(0.025)
Counterfactual mean	0.339	0.407	0.201	0.262	0.870	0.887
Observations	29,496	15,283	29,496	15,283	29,496	15,283

Notes: This table interacts the placement indicator with the educator's narrative recommendation and the formal post-investigation decision in favour of a closed centre, on the full report-linked PARCOURS sample under the entropy-balancing weights, covariates, year and court-department fixed effects of the analysis. The baseline placement coefficient refers to placed minors carrying neither recommendation; the interactions test whether the effect differs for recommended minors, which arbitrates between two readings of the gap between the full-sample and the within-recommended estimates: purged selection or effect heterogeneity. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, in parentheses, are clustered at the court-department level. Source: SSER data [PARCOURS, 2021-2024].

month detention estimate is 11.8 points; adding offence controls raises it slightly to 12.1 points, and balancing on offences raises it to 12.6. These changes are small relative to the standard errors. More importantly, they indicate that if CEF-placed minors were detained simply because they committed more serious offences, this would drive the result and conditioning on offence characteristics should reduce the coefficient. Instead, it increases, which is consistent with the allocation results, where offence characteristics add little to the prediction of CEF placement once the recommendation is observed, and with the DPJJ's description of placement decisions as relying primarily on the minor's broader situation rather than on the offence alone (Direction de la Protection Judiciaire de la Jeunesse, 2021).

Effects on the measures the placement displaces using the historical registries. The recommended population gives the better counterfactual but is limited by a short horizon and a thin sample. The historical registers of GAME/IMAGES bear the opposite limitations but are large enough to estimate the effect against each displaced alternative separately. I therefore estimate the same model of equation (2.4) on the common-support sample with all eligible comparison spells, and read the two sources as complementary rather than interchangeable estimates of one parameter. Panel A of Table 2.12 shows that placement raises the probability of a detention episode within twelve months of the decision by 20.4 percentage points against a counterfactual rate of 21.6%, and by 21.5 points within eighteen months against a rate of 24.9%. Subsequent closed-centre placement rises by 13.0 and 16.0 points and new court referrals by 16.6 and 12.2.

However, this average conceals displacements against different types of measures that can be quite different. Therefore, I re-estimate the same equation against each alternative, holding the treated group fixed. On detention within twelve months of the decision, the effect is 21.5 points against community supervision (Panel B), 15.4 against judicial supervision alone (Panel C), 15.1

Table 2.11: Placement effect on the subsample with a recorded offence nature, outcomes anchored at the decision (PARCOURS, 2021-2024)

	Detention episode		New CEF placement		New court referral	
	12 months	18 months	12 months	18 months	12 months	18 months
Panel A: Covariate adjustment only						
Placed in a closed educational centre	0.146***	0.115***	0.071*	0.117*	0.051***	0.040***
	(0.039)	(0.040)	(0.043)	(0.059)	(0.016)	(0.015)
Counterfactual mean	0.360	0.434	0.212	0.266	0.873	0.887
Observations	15,701	7,455	15,701	7,455	15,701	7,455
Treated observations	191	103	191	103	191	103
Adjusted R^2	0.292	0.389	0.178	0.211	0.150	0.165
Panel B: Adding the recommendation indicators						
Placed in a closed educational centre	0.118***	0.059	0.072	0.116*	0.048***	0.045***
	(0.038)	(0.038)	(0.044)	(0.060)	(0.015)	(0.015)
Counterfactual mean	0.360	0.434	0.212	0.266	0.873	0.887
Observations	15,701	7,455	15,701	7,455	15,701	7,455
Treated observations	191	103	191	103	191	103
Adjusted R^2	0.295	0.395	0.178	0.212	0.150	0.165
Panel C: Adding the offence families as controls						
Placed in a closed educational centre	0.121***	0.061	0.073*	0.122**	0.052***	0.048***
	(0.037)	(0.037)	(0.044)	(0.058)	(0.018)	(0.017)
Counterfactual mean	0.360	0.434	0.212	0.266	0.873	0.887
Observations	15,701	7,455	15,701	7,455	15,701	7,455
Treated observations	191	103	191	103	191	103
Adjusted R^2	0.299	0.398	0.182	0.214	0.159	0.174
Panel D: Reweighting on the offence families as well						
Placed in a closed educational centre	0.126***	0.069*	0.065	0.115**	0.049***	0.048***
	(0.036)	(0.037)	(0.043)	(0.057)	(0.018)	(0.018)
Counterfactual mean	0.356	0.430	0.209	0.258	0.867	0.884
Observations	15,701	7,455	15,701	7,455	15,701	7,455
Treated observations	191	103	191	103	191	103
Adjusted R^2	0.299	0.396	0.181	0.203	0.158	0.174
Included in all regressions:						
Pre-spell judicial history and demographic controls	Y	Y	Y	Y	Y	Y
Year fixed effects	Y	Y	Y	Y	Y	Y
Department fixed effects	Y	Y	Y	Y	Y	Y

Notes: This table reports weighted least squares estimates of the average effect of a first closed educational centre placement on the treated. Each panel restricts the control group to the indicated type of alternative measure; unless stated otherwise the treated group is identical across panels and the entropy-balancing weights are re-estimated within each panel so that the restricted control pool matches the treated covariate profile (weighting method by panel: Covariate adjustment only: Entropy; Adding the recommendation indicators: Entropy; Adding the offence families as controls: Entropy; Reweighting on the offence families as well: Entropy). Each column corresponds to one post-spell outcome and horizon; outcomes at twelve months are defined only for spells whose full one-year post-spell window is observable before the data end and the minor's eighteenth birthday, and therefore over-represent minors placed young. The counterfactual mean is the weighted control mean of the outcome. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, in parentheses, are clustered at the prescriber-department level. The sample covers PARCOURS penal measures with a linked social-educational report and a recorded nature of case; treated units are measures effectively executed in a closed educational centre and controls are all other measures of that sample, so the comparison is not restricted to the recommended population. Outcomes are measured over fixed windows of twelve and eighteen months running from the date of the decision, so that both arms are observed over the same length of time and every subsequent event counts, including a custody that ends the measure while it is still running. A window is defined only where it is fully observable before the end of the data and the minor's eighteenth birthday. These estimates incorporate the revocation channel and answer a different question from the corresponding end-of-measure estimates. The four panels add, in turn, the recommendation indicators, the twelve offence families as regression controls, and the offence families in the balancing set of the weights. Source: SSER data [PARCOURS, 2021-2024].

against a standard residential placement (Panel D), 16.9 against a reinforced educational centre (Panel G) and 11.0 against detention (Panel H).

Because the treated group is held fixed across panels, each estimate changes only the alternative against which the same minors placed in CEF are compared to. Therefore, they ask how their subsequent detention have differed had they received each alternative measure instead. However, the credibility of that comparison differs across alternatives. Against detention, the

estimate remains positive when follow-up starts at the decision date. This differs from the end-of-measure specification displayed in Table F2 because detention following the revocation of a CEF placement is then included in the outcome window, while a minor initially sent to detention spends part of the same period in custody. Against community supervision, which is the main alternative displaced by CEF placement, the estimated increase in subsequent detention is the largest.

Table 2.12: Effect of a first CEF placement by counterfactual type, twelve and eighteen months from the decision

	Detention episode		New CEF placement		New court referral	
	12 months	18 months	12 months	18 months	12 months	18 months
Panel A: All eligible controls						
Placed in a closed educational centre	0.204***	0.215***	0.130***	0.148***	0.166***	0.149***
	(0.010)	(0.011)	(0.011)	(0.013)	(0.008)	(0.009)
Counterfactual mean	0.216	0.249	0.147	0.182	0.484	0.562
Observations	255,021	193,050	255,021	193,050	255,021	193,050
Treated observations	8,381	6,196	8,381	6,196	8,381	6,196
Adjusted R^2	0.195	0.227	0.098	0.123	0.182	0.198
Panel B: Community supervision (all)						
Placed in a closed educational centre	0.215***	0.228***	0.153***	0.173***	0.197***	0.171***
	(0.011)	(0.011)	(0.012)	(0.013)	(0.009)	(0.009)
Counterfactual mean	0.200	0.231	0.122	0.155	0.449	0.537
Observations	185,759	142,784	185,759	142,784	185,759	142,784
Treated observations	8,381	6,196	8,381	6,196	8,381	6,196
Adjusted R^2	0.212	0.233	0.100	0.124	0.176	0.181
Panel C: Judicial supervision only						
Placed in a closed educational centre	0.154***	0.160***	0.092***	0.106***	0.106***	0.077***
	(0.011)	(0.012)	(0.012)	(0.014)	(0.010)	(0.011)
Counterfactual mean	0.267	0.318	0.180	0.225	0.534	0.638
Observations	47,398	34,338	47,398	34,338	47,398	34,338
Treated observations	8,381	6,196	8,381	6,196	8,381	6,196
Adjusted R^2	0.191	0.212	0.110	0.131	0.147	0.159
Panel D: Pre-trial supervised release						
Placed in a closed educational centre	0.251***	0.263***	0.168***	0.180***	0.182***	0.158***
	(0.017)	(0.015)	(0.015)	(0.016)	(0.012)	(0.014)
Counterfactual mean	0.178	0.209	0.109	0.149	0.474	0.561
Observations	78,793	68,057	78,793	68,057	78,793	68,057
Treated observations	8,381	6,196	8,381	6,196	8,381	6,196
Adjusted R^2	0.213	0.249	0.119	0.141	0.179	0.197
Panel E: Suspended sentence / supervised release						
Placed in a closed educational centre	0.210***	0.216***	0.168***	0.195***	0.226***	0.209***
	(0.012)	(0.013)	(0.013)	(0.015)	(0.011)	(0.011)
Counterfactual mean	0.211	0.252	0.112	0.142	0.439	0.523
Observations	38,079	27,625	38,079	27,625	38,079	27,625
Treated observations	8,381	6,196	8,381	6,196	8,381	6,196
Adjusted R^2	0.227	0.269	0.116	0.143	0.182	0.202
Panel F: Residential placement (excl. CER)						
Placed in a closed educational centre	0.151***	0.154***	0.058***	0.064***	0.069***	0.047***
	(0.011)	(0.012)	(0.013)	(0.015)	(0.011)	(0.012)
Counterfactual mean	0.280	0.333	0.231	0.289	0.601	0.698
Observations	38,470	28,214	38,470	28,214	38,470	28,214
Treated observations	8,381	6,196	8,381	6,196	8,381	6,196
Adjusted R^2	0.159	0.189	0.078	0.089	0.100	0.126
Panel G: Reinforced educational centre (CER)						
Placed in a closed educational centre	0.169***	0.154***	0.070***	0.057***	0.116***	0.100***
	(0.013)	(0.015)	(0.016)	(0.021)	(0.010)	(0.011)
Counterfactual mean	0.259	0.324	0.213	0.296	0.553	0.639
Observations	15,439	11,200	15,439	11,200	15,439	11,200
Treated observations	8,381	6,196	8,381	6,196	8,381	6,196
Adjusted R^2	0.169	0.195	0.071	0.077	0.142	0.189
Panel H: Detention						
Placed in a closed educational centre	0.110***	0.097***	0.018	0.036*	-0.053***	-0.032**
	(0.015)	(0.016)	(0.015)	(0.019)	(0.013)	(0.013)
Counterfactual mean	0.319	0.377	0.227	0.254	0.716	0.756

Table 2.12: Effect of a first CEF placement by counterfactual type, twelve and eighteen months from the decision *(continued)*

	12 months	18 months	12 months	18 months	12 months	18 months
Observations	20,444	13,217	20,444	13,217	20,444	13,217
Treated observations	8,381	6,196	8,381	6,196	8,381	6,196
Adjusted R^2	0.163	0.202	0.101	0.126	0.098	0.140
Included in all regressions:						
Pre-spell judicial history and demographic controls	Y	Y	Y	Y	Y	Y
Year fixed effects	Y	Y	Y	Y	Y	Y
Department fixed effects	Y	Y	Y	Y	Y	Y

Notes: This table reports weighted least squares estimates of the average effect of a first closed educational centre placement on the treated. Each panel restricts the control group to the indicated type of alternative measure; unless stated otherwise the treated group is identical across panels and the entropy-balancing weights are re-estimated within each panel so that the restricted control pool matches the treated covariate profile (weighting method by panel: All eligible controls: Entropy; Community supervision (all): Entropy; Judicial supervision only: Entropy; Pre-trial supervised release: IPW; Suspended sentence / supervised release: Entropy; Residential placement (excl. CER): Entropy; Reinforced educational centre (CER): Entropy; Detention: Entropy). Each column corresponds to one post-spell outcome and horizon; outcomes at twelve months are defined only for spells whose full one-year post-spell window is observable before the data end and the minor's eighteenth birthday, and therefore over-represent minors placed young. The counterfactual mean is the weighted control mean of the outcome. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, in parentheses, are clustered at the prescriber-department level. Outcomes are anchored at the decision date of the indexed spell over fixed windows of twelve and eighteen months, identical on both arms, and count every subsequent event including a detention that revokes the measure under way. Eighteen months is chosen because a placement lasts about six months, so the window leaves roughly a year beyond the end of a typical stay while keeping the length of observation the same for placed and comparison minors; the end-of-measure anchoring used elsewhere cannot do this, since measures end at systematically different ages on the two arms. These estimates therefore incorporate the revocation channel and answer a different question from the end-of-measure estimates reported in the appendix. Source: SSER data [GAME/IMAGES, 2012-2022].

The same legal vehicle, with and without the placement. A closed centre placement is never ordered on its own, it is attached to a judicial supervision order or to a suspended sentence, which is the legal device that makes the closed regime enforceable. The right question is therefore what the placement adds to that order, and the comparison group is minors carrying the same order without the placement. Table E2 of Appendix E reports this. Adding a placement to a judicial supervision order raises the probability of a detention episode within twelve months of the decision by 15.9 points on a base of 30.7%, and adding it to a suspended sentence by 20.9 points on a base of 39.5%. Both arms are held under the same coercive instrument and differ only in whether the minor is placed in a closed centre. Therefore the difference keeps the underlying legal vehicle constant and is only driven by the residential component of the CEF. The same comparison can be run on the recommended population sample, where the educator's recommendation is observed and held constant. Restricting the controls of Table 2.8 to recommended minors who received a judicial supervision order instead of the placement returns 5.5 points on detention within 12 months of the decision, against 8 points for the full recommended sample at the same horizon, on 660 observations. The estimate is smaller than on the registry, as every estimate on the recommended population is, but the sign and the order of magnitude survive.

Sensitivity to selection and to the adjustment. The design rests on selection on observables, and I therefore run sensitivity analyses using two families of statistics to quantify how strong unobserved selection would have to be to overturn the estimated effects. They are computed on the set of covariates of the main specification with outcomes starting at the decision of placement. I apply the coefficient-stability logic of Oster (2019), and find that selection on unobservables would have to be 4.0 times as strong as selection on the twenty-two observed covariates to drive the 12-months detention estimate on the historical registers to zero, and 3.1 times at 18 months (Table C3). That statistic is indexed to a proportionality convention rather than to a quantity the data identify, so I complement it with the partial R^2 approach of Cinelli and Hazlett (2020), which asks how much of the residual variation in the treatment and in the outcome an unobserved confounder would have to explain, *jointly*, to bring the estimate to zero. Table C4 reports both families of statistics on two samples. On the report-linked sample, where the recommendation varies and enters as a control, the coefficient-stability threshold is 1.9 at 12 months and 0.9 at 18, and a single unrecorded factor would have to explain 4.4 and 2.8% of that residual variation to bring the coefficient to zero. The natural benchmark is the educator's recommendation itself, since it is the closest available summary of the case-file information the administrative records omit. Those figures are not reassuring at the longer horizon, where the estimate is small and the fixed window halves the sample.

On the recommended population, which is the sample of the main estimate of the paper, both families point the other way. At twelve months the coefficient-stability threshold is negative, which indicates that the estimate moves *away* from zero as the covariates are added: under selection on unrecorded factors as strong as selection on the recorded ones, the coefficient rises to 0.092 from an adjusted 0.086. Proportional selection of that kind would increase the estimate rather than drive it to zero. At 18 months, the threshold is 5.8. The share of residual variation a confounder would have to carry on both placement and detentino at once is 7.1% at 12 months and 4.0% at 18, against 4.4 and 2.8% on the report-linked sample.

In order to grasp interpret these statistics, I compare them to what the recorded information itself explains. On the recommended population at 12 months, the whole set of pre-measure covariates accounts for 6.3% of the residual variation in placement and 16.4% of the residual variation in detention. The threshold of 7.1% therefore describes a factor that would have to be, *on its own*, slightly more predictive of the placement than the twenty-two recorded covariates taken together, and *simultaneously* as predictive of the outcome. On the report-linked sample, the educator's recommendation can serve as a second comparison point, since it is available there: it accounts for 5.7% of the residual variation in placement and 0.3% in detention, and an unrecorded factor would have to be 2.9 times as predictive as the recommendation on *both* margins at once to bring the 12-months coefficient to zero. Conditioning on the institution's own assessment makes the estimate more robust to unobserved selection, therefore strengthening the design.

Three further checks bear on the adjustment rather than on the unobserved. First, I run a falsification test on pre-determined variables deliberately excluded from the balancing set. Table C5 shows that it partly fails, since placed minors come from a more recent judicial contact

by 71 days and carry 5.3 points more pre-decision detention and 9.0 points more pre-decision court referral than their weighted controls. However, adding all of these to the conditioning set moves the 12-months estimate from 20.8 to 20.3 points and the 18-months one from 22.0 to 21.5 (Table C6): the imbalance is real, and correcting it leaves the estimate where it was.

Additionally, I show in Table C7 where the adjustment comes from by presenting estimates from a range specifications: the raw contrast with only year and prescriber-department fixed effects goes from 32.1 points to 19.8 when adding covariates alone, to 24.2 with inverse-probability weights (IPW) alone instead of the covariates and 20.4 with both. I run the same specifications with entropy balancing weights (EB), and find that, reassuringly, the result is not specific to the weighting method as IPW and entropy balancing agree to within one tenth of a point.

Finally, Table C8 shows that the support rule does not drive the result either: the tenth-percentile threshold retains 90% of treated placements, while the twenty-fifth percentile drops around a quarter of them for only a modest improvement in overlap, despite the fact that the estimates are essentially unchanged. I therefore use the tenth-percentile rule in the main specification, which improves overlap while retaining substantially more of the treated population.

Robustness. I address the main concerns surrounding this selection-on-observables design, and the estimates remain stable across all of them. The first concerns whether the identification comes from differences across jurisdictions. Table C9 shows that absorbing the deciding court instead of the prescriber department lowers the twelve-month estimate on the recommended population from 8.0 to 6.7 points, and absorbing the court interacted with the year lowers it to 6.1, significant at the 5% level in both cases and on exactly the same sample. Therefore the effect does not appear to come from comparing the effect is not a comparison between jurisdictions that place readily and jurisdictions that do not.

The second concerns the unit. Since a minor with a long trajectory supplies more comparison decisions than another: giving every untreated minor the same total base weight within each comparison, and re-imposing exact balance on the down-weighted pool, raises the twelve-month detention estimate on the registers from 20.4 to 23.2 points and leaves every panel positive and significant (Table C10). The effects on a new closed centre placement and on a new court referral are more sensitive to this weighting than the detention estimates are.

The third concerns the balancing set: Table C11 shows that moving the pre-decision variables into it from the start, rather than adding them afterwards as a correction, changes the estimates by less than one percentage point in every panel, with no change of sign or of significance anywhere.

The fourth concerns the selection into the report-linked sample used for the recommendation-based estimate. Report linkage is systematically related to the court, the year and the observed characteristics of the case. However, I show in Table B7 that it does not predict subsequent detention among closed centre placements once those dimensions are absorbed: the conditional difference is 1.9 percentage points at 12 months (standard error 2.2) and 1.1 points at 18 months (standard error 3.7), on mean detention rates of 45 and 53%. Linked placements do show a 6.2-point higher 12-month court-referral rate, so linkage is not treated as random.

Additionally, entropy balancing attains covariate balance by weighting, and the concentration of these weights varies substantially across comparisons. On the pooled comparison the reweighting is well spread, with an effective sample size of 165,709 for 339,622 comparison spells and 7.3% of the total weight carried by the heaviest 1%. The comparison with pre-trial supervised release is much more concentrated: the effective sample size falls to 6,435 for 73,958 comparison spells, and the heaviest 1% carries 27.1% of the weight (Table C12). That panel therefore rests on far fewer effective comparisons than its sample size suggests, and its coefficient, which is among the largest across alternatives, should be read with that in mind. In the seven remaining comparisons, the heaviest 1% carries between 4 and 11 per cent of the total weight.

The outcome anchor explains the difference between the estimates reported here and those in Appendix F. Measuring the same outcome from the end of the measure rather than from the decision, which removes from the count every custody episode beginning through the revocation of an ongoing placement, lowers the 12-months detention estimate on the registers from 20.4 to 13.6 points (Table F2). This difference is consistent with revocation playing a particular role after CEF placement: among the detentions that follow within 12 months, the share beginning within seven days of the day the measure ends reaches 26.8% after a closed centre placement against 16.8% after another residential placement, 11.3% after a reinforced centre and 7.1% after a standalone supervision order (Table G1 of Appendix G). The comparison pool provides a second robustness check. The main pool drops placements that are the only judicial record a minor has, on the ground that a first and only appearance carries no trajectory to match on; putting them back lowers the 12-month estimate from 20.4 to 18.6 points and leaves every sign and significance level unchanged (Table E3). Finally, the support rule is estimated once against all alternatives, the threshold applied to any one alternative is partly determined by observations receiving other measures; re-estimating the score and the support within each comparison separately leaves the pattern of estimates essentially unchanged (Table E1).

6 Mechanisms

I find that the placement raises the probability of subsequent custody, but the programme could plausibly harm through the experience it delivers, through the company it imposes, or through the legal device that carries it, and these have different policy implications. This section examines the channels the data allow: the judicialisation of failure during the stay, the legal vehicle, the regime inside the facility, and the programme actually completed.

Failure during the stay and its judicialisation. If the placement worked as a treatment whose consequences appear afterwards, the custody that follows it would occur after the minor leaves. The timing can be checked directly. Among first placements whose 24-month window is fully observable, 56.4% are followed by a detention episode within 2 years of the decision, and 43.0% of those first detentions begin while the placement measure is still formally running. Table 2.13 reports the distribution of the delay between the end of the measure and the first detention, and crossing the detention dates with the monthly housing records shows that 26.4%

of first detentions start in the same month as a recorded runaway or in the following one, rising to 34% among detentions occurring during the spell.

Table 2.13: Delay between the end of the indexed spell and the first subsequent detention episode

Delay from spell end to first detention	Eligible controls	Placed in a CEF
0-30 days	12.1%	20.0%
31-90 days	20.3%	20.7%
91-365 days	46.6%	43.9%
Over one year	21.1%	15.5%

Notes: This table distributes, among spells followed by at least one detention episode before age eighteen, the delay between the recorded end of the indexed spell and the start of the first detention episode, by arm, in the common-support sample drawn from the GAME/IMAGES records. A concentration in the first month is consistent with quasi-mechanical transitions, namely revocation of the underlying order after a failed measure, or a detention that itself terminates the spell. The distribution is descriptive; it conditions on an outcome. Source: SSER data [GAME/IMAGES, 2012-2022].

Recorded runaways understate the number of incidents, since short absences and disciplinary exclusions leave no trace, so the coincidence documented here is itself understated. The timing and the institutional asymmetry are consistent with revocation being an important channel: a substantial share of the custody that follows a placement begins while the measure is still running and close to a recorded absence, which is what a breach of the supervision order converting into custody would produce. What these figures do not identify is the share of the estimated effect that travels through that channel, since the comparison is descriptive of the placed arm rather than of a mediation. These episodes are counted by the main estimates, whose window opens at the decision, and are excluded by construction from the end-of-measure specification reported in Appendix F. The difference between the two anchorings measures, as a matter of accounting on the same sample, what the custody occurring during the measure contributes to the estimate: on the registers it is the gap between 20.4 and 13.6 points at twelve months.

The legal vehicle. If the placement stabilised the judicial relationship, supervision orders carrying a CEF placement would run longer than those without. Figure G1 in Appendix G tests this by comparing, through Kaplan-Meier survival curves, supervision orders started between 2012 and 2022 with and without a placement attached within seven days of their start, excluding orders that receive a placement later on. This exclusion matters as an order that receives a placement after six months had to survive those six months to be classified as placed, so keeping such orders in the placed group would make placement appear to lengthen orders purely by construction. The median duration is 353 days with a placement against 532 without, and at one year 47% are still running against 63%. The placement does not extend judicial retention but precipitates its rupture, which is the duration counterpart of the revocation channel documented above. Minors whose order carries a placement are not randomly selected, so the comparison is descriptive.

The regime inside the facility. Whether the closed centre is more disruptive than the residential alternative the same minor would otherwise have received is not observable in the main analytical sample, so I draw on the monthly housing records, which report the days spent in each housing structure and the monthly number of days that the minor lost to runaway, to incarceration during the day, to hospitalisation and to administrative interruptions. The comparison uses the same reweighted residential counterfactual as the effect estimates, so the mechanism evidence rests on the design of Section 4.4 rather than on a separate one. Table 2.14 reports the gaps.

Table 2.14: In-stay incidents: closed educational centre against the reweighted residential counterfactual

	Substantive incidents				Regime of the placement	
	Ran away	Incarcerated during stay	Hospitalised during stay	Any substantive incident	Authorised leave	Administrative interruption
Placed in a closed educational centre	0.102***	0.115***	0.007***	0.146***	-0.063***	-0.000
	(0.014)	(0.008)	(0.003)	(0.014)	(0.015)	(0.010)
Counterfactual mean	0.359	0.079	0.018	0.405	0.333	0.095
Observations	33,744	33,744	33,744	33,744	33,744	33,744
Included in all regressions:						
Pre-spell controls	Y	Y	Y	Y	Y	Y
Year fixed effects	Y	Y	Y	Y	Y	Y
Department fixed effects	Y	Y	Y	Y	Y	Y

Notes: This table compares closed educational centre placements with the reweighted residential counterfactual on events recorded during the stay, using the monthly housing records matched to the indexed spell by minor, taking-charge identifier and month. Each column is one indicator equal to one when at least one day is recorded in the corresponding category; estimates come from weighted least squares with the entropy-balancing weights, pre-spell controls, year and prescriber-department fixed effects. The last two columns describe the closed regime rather than incidents: authorised leave is by design rarer in a closed facility, so the negative gap measures the restriction of movement that defines the placement and cannot serve as a falsification test. Housing coverage is partial and selective on observables. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, in parentheses, are clustered at the prescriber-department level. Source: SSER data [GAME/IMAGES-HEBERGEMENT, 2012-2022].

I find that closed-centre stays record 10.2 additional percentage points of runaways, 46 against 36%, and 11.5 additional points of within-stay incarceration, 19 against 8%, and 14.6 additional points of any substantive incident. Administrative interruptions show no gap, which serves as an internal control since a purely selection-driven story would disturb the administrative profile as well. Authorised leave is 6.3 points rarer, but this measures the restriction of movement that defines a closed facility rather than an incident.

The timing of these incidents, presented in Figure 2.2, is informative about the sequence. Runaways rise immediately, peak in the second month at around 4 days per spell-month and remain high through the sixth. Within-stay incarcerations follow with a $1-2$ month lag and becomes stationary around 1.5 days per spell-month. The lag is what the revocation channel predicts, since an incident converts into custody only after a judicial decision that takes weeks. By the fourth month, when the reintegration phase is supposed to begin, 45% of stays have already ended.

How much of the effect travels through these incidents can then be estimated within the treated arm. Table 2.15 regresses post-placement detention on the joint distribution of runaway and within-stay incarceration, with stays free of both as the reference. A stay with a runaway

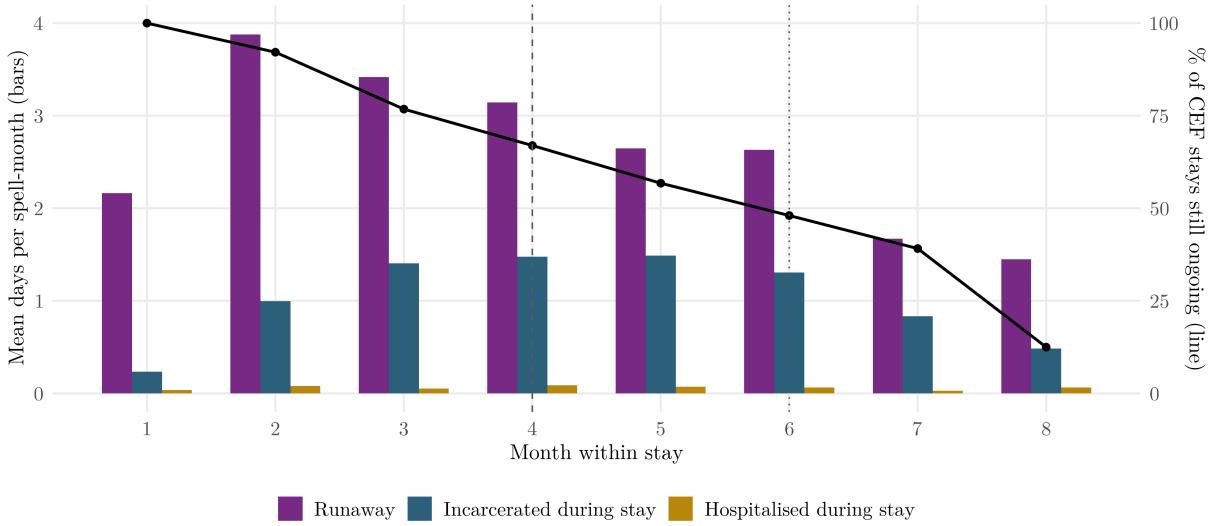

Figure 2.2: Within-stay events and exit pattern of the closed educational centre stay

Notes: This figure plots the within-stay profile of incidents (left axis, bars) and the exit pattern of the stay (right axis, line) against the month elapsed since the start of the indexed placement, on the treated arm of the common-support sample observed in the monthly housing records. Bars report, for each month, the mean number of days minors spent on the run from the placement, incarcerated during the stay or hospitalised, computed only over stays still observed at the indicated month. The line reports the share of stays still observed at each month relative to the first month. The dashed line at four months marks the start of the final reintegration phase of the programme; the dotted line at six months marks the statutory upper bound, beyond which stays correspond to renewals. Source: SSER data [GAME/IMAGES-HEBERGEMENT, 2012-2022].

but no within-stay incarceration carries around 15 additional points of post-placement detention, one with a within-stay incarceration carries between 13 and 15 percentage points depending on the horizon of the outcome considered.

The interaction is negative, so the two channels do not compound. Both effects are precisely estimated and persist across horizons, but they cannot account for the whole of the effect: the remainder must pass through the experience of the stay itself, through judicial path dependence once a minor carries a closed-centre record, and through the return to the home environment.

These comparisons are descriptive of the institutional regime under an observable-balance assumption. The design equalises the recorded trajectory but cannot speak to selection on dimensions the source does not record, so the contrast may combine a regime effect, the programme producing more incidents by its structure, with a residual selection effect, the programme admitting minors who would have had more incidents in any residential setting. Housing coverage is itself partial and selective on observables, so the gaps should be read as suggestive evidence.

Table 2.15: Decomposition of post-spell detention by within-stay incident profile

	Detention within 12 months		Detention within 18 months	
	(1)	(2)	(1)	(2)
Within-stay runaway	0.179***	0.153***	0.188***	0.167***
	(0.024)	(0.022)	(0.030)	(0.027)
Within-stay incarceration	0.184***	0.133***	0.292***	0.247***
	(0.044)	(0.040)	(0.041)	(0.041)
Runaway × incarceration	-0.162***	-0.127**	-0.200***	-0.182***
	(0.056)	(0.053)	(0.055)	(0.054)
No-incident mean	0.341	0.341	0.409	0.409
Pre-spell controls	N	Y	N	Y
Year fixed effects	Y	Y	Y	Y
Department fixed effects	Y	Y	Y	Y
Observations	2,637	2,637	1,754	1,754
Adjusted R^2	0.076	0.105	0.131	0.174

Notes: This table restricts the sample to closed educational centre placements observed in the monthly housing records and regresses post-spell detention on indicators of a runaway and of an incarceration occurring during the stay, and on their interaction, with the no-incident group as reference. Columns (1) report the bare specification and columns (2) add the pre-spell controls of the pre-spell judicial history and demographics; all specifications include year and department fixed effects. Outcome windows open at the end of the measure and close after twelve and eighteen months. A within-stay incarceration can extend past the administrative end of the placement and mechanically produce a recorded post-spell detention, whereas a runaway marks a destabilised stay without that mechanical link. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, in parentheses, are clustered at the prescriber-department level. Source: SSER data [GAME/IMAGES-HEBERGEMENT, 2012-2022].

The programme actually delivered. The three-phase programme described in Section 2.1 presupposes that the minor goes through all of it, and the average effective stay of 4.5 months with a standard deviation of about three implies that a substantial share leaves before the four-month point at which the exit-preparation phase begins.

Figure 2.3 splits the treated arm at four months and rebuilds a counterfactual for each group, the residential comparison pool being reweighted separately against each group's covariate profile. The adverse pattern is concentrated on stays cut short, while completed stays track their counterfactual far more closely. The same split, estimated on the recommended population, returns 20.7 points on detention within twelve months of the decision for stays ending before four months and -6.2 points, statistically indistinguishable from zero, for stays of four months or more (Table G2 of Appendix G). The gap between the two panels is a twenty-seven point spread, and it widens rather than narrows when the window opens at the decision, since the custody that cuts a stay short is then counted rather than removed; the end-of-measure reading of the same split is reported in Table F3 of Appendix F. Effective stay length is itself an outcome of the placement rather than a randomly assigned dimension of the treatment, so the contrast is suggestive only. Bruyn et al. (2012) documents that premature departure is rarely a planned exit and is instead associated with incidents during the stay, with prior incarcerations, and with supervision revocations leading to emergency incarceration; minors who leave before four months are therefore not a random sub-sample of the treated arm, and the contrast mixes a programme-completion effect with selection on the trajectory of the stay.

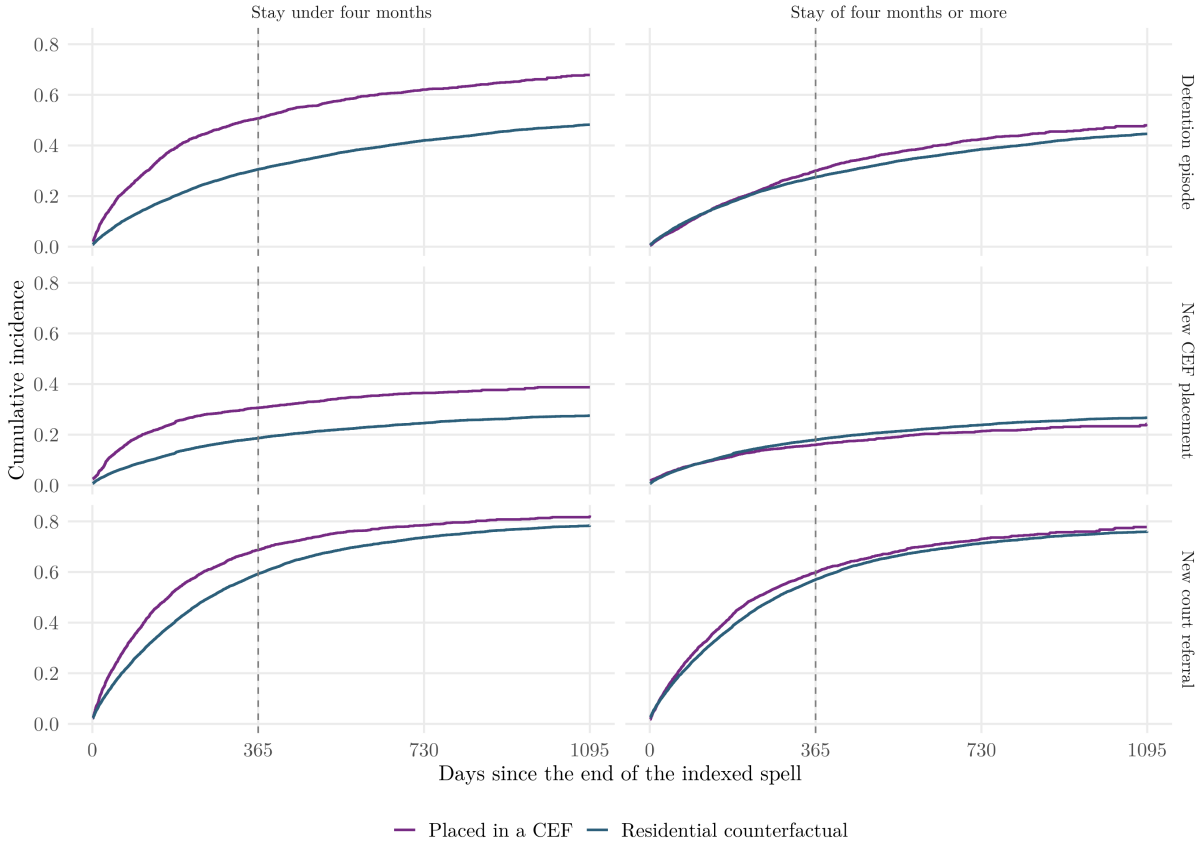

Figure 2.3: Cumulative incidence of post-spell events by effective stay length

Notes: This figure plots cumulative incidence curves of the time to the first subsequent event, computed as one minus the weighted Kaplan-Meier survival estimate and right-censored at the maximum observable post-spell window. The left column restricts the treated arm to closed educational centre stays shorter than four months, the right column to stays of four months or more; in each column the residential control pool is reweighted by entropy balancing against the covariate profile of that treated group, so each column has its own counterfactual. Each row tracks one event type. The dashed line marks the twelve-month horizon and the window is truncated at three years. Effective stay length is an outcome of the placement rather than an assigned dose. Source: SSER data [GAME/IMAGES-HEBERGEMENT, 2012-2022].

Whether what matters is the operator or the programme it delivers can be tested on the same sample. Publicly run centres record 4.7 additional points of twelve-month detention relative to accredited associative ones, which suggests a sector effect. Table G3 of Appendix G adds interactions with terciles of the facility's completion rate, the share of its terminated stays reaching six months. The public differential on detention then falls to 1.4 points and loses significance while the completion terciles absorb the gap, so what matters is the programme actually delivered rather than the status of the operator. This is consistent with the earlier finding that the severity of the profile divides facilities without dividing sectors. The residual public differential on new court referrals, around 4 percentage points, is not absorbed and remains unexplained. Because publicly placed minors carry observably heavier histories and the weights hold that composition constant without excluding its unobservable counterpart, the sector gap should be read as an upper bound.

Two features of the material conditions complete the picture. Occupancy computed in occu-

pied days over capacity-days reaches 103 to 106% of authorised capacity in the administrative records, while direct housing records place physical presence around ten minors for twelve places, and 16.5% of measure-days are absence days. The gap between the two counts seems to be due to beds that are occupied on paper and empty in fact because the minor has run away or been incarcerated, the measure stays open, and the place stays assigned to them whether or not it is used in the meantime. An occupancy rate computed from the administrative records therefore stands above the number of minors physically present, which is why the two sources disagree.

Table G4 of Appendix G reports the effect separately according to whether the receiving centre lies in the prescriber's department, in another department of the same regional directorate, or in another directorate altogether. The 77.8% of placements executed outside the prescribing department are not in tension with the proximity finding of Section 5.1, since most departments operate no centre at all, so even the nearest one usually lies across a departmental boundary.

Taken together, these results are consistent with revocation being an important channel of the adverse effect. Incidents are more frequent under the closed regime than under the residential alternative, the custody that follows often begins while the measure is still running, and most minors leave before the final phase of the programme. None of these comparisons identifies the share of the effect that passes through the channel, since each rests on the placed arm or on an observable-balance assumption rather than on assigned variation. The reading they support is institutional rather than behavioural, and it fits the estimate being smallest on the population the institution itself judged suitable for the measure.

One reason a judge may place a minor who does not need containment is that the centre is where the resources are.A closed centre runs on 26.5 full-time-equivalent posts, 17 of them specialised educators, for around 10 minors physically present, while an educator in a community supervision service carries 25 measures.[12] A judge facing a minor who needs psychological care or intensive educational support, in a department where the open-setting service can no longer provide either, may reasonably expect the centre to. I have no budget series showing one funded at the expense of the other, so the staffing ratio is the only evidence I can bring, and I give this as a channel the results are consistent with rather than as a finding. This holds even granting that placed minors are selected on characteristics I do not observe: the revocation mechanism is attached to the measure and not to the minor, so whatever unobserved severity placed minors carry, the comparison group carries no equivalent device.

7 Conclusion

Closed educational centres were created as an alternative to incarceration, but whether they function as one depends on what they replace. The results of this paper show that, for most

[12]The staffing target comes from the 2016 circular (Direction de la Protection Judiciaire de la Jeunesse, 2016) and the count of minors physically present from the monthly housing records. The community supervision figure is reported by Inspection générale de la justice (2025) as a national average for 2024, with 31 measures per educator in the Île-de-France and overseas interregional directorate and 28 in the South-East; the same report uses one educator per 25 minors when costing the reinforcement of community supervision on exit from a closed centre. The report gives no source for either figure, and a measure is not a minor, so the two ratios indicate an order of magnitude rather than a directly comparable ratio.

of the minors who receive a CEF placement, they replace community supervision rather than detention. Therefore, CEF placement usually intensifies an existing judicial response rather than diverting a minor from custody. Among minors whom the institution recommended for a CEF, placement raises detention within 12 months by 8 percentage points. This effect is largest relative to community supervision and remains positive relative to detention itself. The timing of subsequent custody is consistent with revocation of the order carrying the placement being an important part of this effect.

These findings show that an intermediate juvenile measure cannot be understood solely from its position in the gradient of severity measures available to the judge. A measure located immediately below prison can still increase the intensity of intervention if most of the individuals it reaches would otherwise have remained under community supervision. Its legal architecture also matters independently of the content of the programme: when placement is attached to an order that can be revoked after a breach, the legal structure of the measure itself can lead to detention.

This does not imply that the educational programme itself is not effective. The analysis observes judicial trajectories before age 18, not schooling, employment or health, and compares assignment to a CEF with the measure it displaces rather than completed with interrupted placements. The causal interpreation also remains conditional on selection on observables, although the results are stable across the sensitivity analyses considered in the paper.

The distinction between the programme and its legal vehicle has become directly policy-relevant. From 2026, the reform of public-sector CEFs removes the possibility of revoking judicial supervision for part of their target population, while the associative sector initially keeps it. If revocation contributes to the detention effect documented here, subsequent cohorts provide a natural setting in which to test whether changing the legal architecture changes later custody. Linking these records to education, employment and health data would also allow the rehabilitative effects of the programme to be assessed.

References

Abrams, David, Roberto Galbiati, Emeric Henry, and Arnaud Philippe (2022). "When in Rome... On Local Norms and Sentencing Decisions". *Journal of the European Economic Association* 20.2, pp. 1234–1274 (cit. on p. 75).

Agan, Amanda, Jennifer L. Doleac, and Anna Harvey (2023). "Misdemeanor Prosecution". *Quarterly Journal of Economics* 138.3, pp. 1453–1505 (cit. on p. 76).

Aizer, Anna and Joseph J. Doyle Jr. (2015). "Juvenile Incarceration, Human Capital, and Future Crime: Evidence from Randomly Assigned Judges". *Quarterly Journal of Economics* 130.2, pp. 759–803 (cit. on p. 73).

Angelova, Victoria, Will Dobbie, and Crystal S. Yang (2026). "Algorithmic Recommendations and Human Discretion". *Review of Economic Studies* 93, pp. 2250–2283 (cit. on p. 75).

Arbour, William (2022). "Can Recidivism Be Prevented From Behind Bars? Evidence From a Behavioral Program" (cit. on p. 74).

Arbour, William and Steeve Marchand (2025). "Can Parole Reduce Both Time Served and Crime?" (Cit. on p. 75).

Baron, E. Jason, Brian Jacob, and Joseph Ryan (2023). "Pretrial juvenile detention". *Journal of Public Economics* 217, p. 104798 (cit. on p. 73).

Bayer, Patrick, Randi Hjalmarsson, and David Pozen (2009). "Building Criminal Capital Behind Bars: Peer Effects in Juvenile Corrections". *Quarterly Journal of Economics* 124.1, pp. 105–147 (cit. on pp. 74, 99).

Bhuller, Manudeep, Gordon B. Dahl, Katrine V. Løken, and Magne Mogstad (2020). "Incarceration, Recidivism, and Employment". *Journal of Political Economy* 128.4, pp. 1269–1324 (cit. on p. 74).

Bhuller, Manudeep, Laura Khoury, and Katrine V. Løken (2025). "Mental Health Consequences of Correctional Sentencing". *American Economic Journal: Economic Policy* 17.1, pp. 70–105 (cit. on p. 74).

Bruyn, Florence de, Luc-Henry Choquet, and Lydia Thierus (2012). "Les "sorties" des mineurs de la délinquance à l'issue d'un séjour en centre éducatif fermé". *Les sorties de délinquance*. Ed. by Marwan Mohammed. Recherches. Paris: La Découverte, pp. 279–298 (cit. on pp. 86, 114).

Carlos, Rita (2020). "Espaces autorisés et autorité des places : des déplacements aux trajectoires des acteurs en Centre éducatif fermé". *Champ pénal* 20 (cit. on p. 78).

Cinelli, Carlos and Chad Hazlett (2020). "Making Sense of Sensitivity: Extending Omitted Variable Bias". *Journal of the Royal Statistical Society Series B* 82.1, pp. 39–67 (cit. on p. 108).

CNAPE (2012). *Les CER : une réponse éducative pertinente qui reste trop méconnue*. Tech. rep. Convention nationale des associations de protection de l'enfant (cit. on p. 80).

Cour des comptes (2023). *Les centres éducatifs fermés et les établissements pénitentiaires pour mineurs*. Tech. rep. S2023-0733. Cour des comptes (cit. on pp. 74, 79–81, 88, 121, 126).

Depew, Briggs, Ozkan Eren, and Naci Mocan (2017). "Judges, Juveniles, and In-Group Bias". *Journal of Law and Economics* 60.2, pp. 209–239 (cit. on p. 75).

Direction de la Protection Judiciaire de la Jeunesse (2016). *Circulaire d'application du 10 mars 2016 de l'arrêté n° JUSF1509326A du 31 mars 2015 relatif aux règles d'organisation, de fonctionnement et de prise en charge des centres éducatifs fermés du secteur public de la protection judiciaire de la jeunesse*. Circulaire NOR JUSF1607483C. Paris: Ministère de la Justice, Direction de la Protection Judiciaire de la Jeunesse (cit. on pp. 74, 78, 79, 81, 116).

— (2020). *Étude relative aux mineurs déférés*. Rapport d'étude. Paris: Ministère de la Justice, DPJJ/SDMPJE/SERC, p. 61 (cit. on pp. 81, 82, 97).

— (2021). *Étude sur le profil des mineurs placés en CEF au 15 juin 2021*. Rapport. Paris: Ministère de la Justice, Sous-direction des missions de protection judiciaire et d'éducation, Service de l'évaluation, de la recherche et du contrôle (cit. on pp. 74, 78, 80–82, 86, 104).

— (2024). *Étude de cohorte sur les parcours de prise en charge à la protection judiciaire de la jeunesse*. Note statistique. Paris: Ministère de la Justice, DPJJ – Service des études, de la recherche et des évaluations et Bureau des systèmes d'information et du contrôle de gestion (cit. on p. 121).

Drago, Francesco and Roberto Galbiati (2012). "Indirect Effects of a Policy Altering Criminal Behavior: Evidence from the Italian Prison Experience". *American Economic Journal: Applied Economics* 4.2, pp. 199–218 (cit. on p. 74).

Drago, Francesco, Roberto Galbiati, and Pietro Vertova (2011). "Prison Conditions and Recidivism". *American Law and Economics Review* 13.1, pp. 103–130 (cit. on p. 74).

Eren, Ozkan and Naci Mocan (2021). “Juvenile Punishment, High School Graduation, and Adult Crime: Evidence from Idiosyncratic Judge Harshness”. *Review of Economics and Statistics* 103.1, pp. 34–47 (cit. on p. 74).

Farcy-Callon, Léo (2016). “La violence en centre éducatif fermé : enjeux et traitement institutionnel”. *Le Sociographe* 56.4, pp. 11–24 (cit. on p. 80).

Fédération Wallonie-Bruxelles (2026). *Les Institutions Publiques de Protection de la Jeunesse (IPPJ)*. Portail de l’Aide à la jeunesse (cit. on p. 73).

Franco, Catalina, David J. Harding, Shawn D. Bushway, and Jeffrey Morenoff (2024). “Does Re-Imprisonment for Technical Violations Prevent Crime?” *Journal of Human Resources* (cit. on p. 75).

Hainmueller, Jens (2012). “Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies”. *Political Analysis* 20.1, pp. 25–46 (cit. on p. 93).

Harrington, Emma, William Murdock III, and Hannah Shaffer (2026). “Prediction Errors, Incarceration, and Violent Crime: Evidence from Linking Prosecutor Surveys to Court Records”. *American Economic Journal: Economic Policy* 18.2, pp. 375–403 (cit. on p. 75).

Heller, Sara B., Anuj K. Shah, Jonathan Guryan, Jens Ludwig, Sendhil Mullainathan, and Harold A. Pollack (2017). “Thinking, Fast and Slow? Some Field Experiments to Reduce Crime and Dropout in Chicago”. *Quarterly Journal of Economics* 132.1, pp. 1–54 (cit. on p. 74).

Hjalmarsson, Randi (2009). “Juvenile Jails: A Path to the Straight and Narrow or to Hardened Criminality?” *Journal of Law and Economics* 52.4, pp. 779–809 (cit. on p. 74).

IGSJ and IGAS (2015). *Rapport sur le dispositif des centres éducatifs fermés (CEF) – Mission d’évaluation du dispositif des centres éducatifs fermés*. Rapport interministériel IGSJ N° 34-15 / IGAS N° 2015-014R. Paris: Inspection générale des services judiciaires (IGSJ), Inspection générale des affaires sociales (IGAS) et Inspection de la protection judiciaire de la jeunesse (IPJJ) (cit. on pp. 74, 78, 79, 81).

Inspection générale de la justice (2025). *Mission thématique sur les Centres Éducatifs Fermés : Rapport définitif*. Tech. rep. N°021 - 2024/00151. Inspection générale de la justice, Ministère de la Justice (cit. on pp. 78, 81, 82, 116).

Lotti, Giulia (2022). “Tough on Young Offenders: Harmful or Helpful?” *Journal of Human Resources* (cit. on p. 74).

Mastrobuoni, Giovanni and Daniele Terlizzese (2022). “Leave the Door Open? Prison Conditions and Recidivism”. *American Economic Journal: Applied Economics* 14.4, pp. 200–233 (cit. on p. 74).

Michel, Bastien and Camille Hémet (2022). “Incarceration versus Probation? Long-Run Evidence from an Anticipated Reform”. *CEPR Discussion Paper No. 15047* (cit. on p. 74).

Ministère de la Justice (2026). *Circulaire de politique pénale et éducative relative à la justice des mineurs*. Circulaire NOR JUSF2604225C. Paris: Ministère de la Justice, Garde des sceaux (cit. on p. 81).

Ministero della Giustizia (2026). *Processo al minore* (cit. on p. 73).

Mucchielli, Laurent (2005). “Les « centres éducatifs fermés » : rupture ou continuité dans le traitement des mineurs délinquants ?” *Revue d’histoire de l’enfance « irrégulière »* 7, pp. 113–146 (cit. on pp. 78, 121).

Mueller-Smith, Michael and Kevin T. Schnepel (2021). “Diversion in the Criminal Justice System”. *Review of Economic Studies* 88.2, pp. 883–936 (cit. on p. 76).

Oster, Emily (2019). “Unobservable Selection and Coefficient Stability: Theory and Evidence”. *Journal of Business & Economic Statistics* 37.2, pp. 187–204 (cit. on pp. 77, 108).

Ouss, Aurélie and Arnaud Philippe (2025). “Prison Peer Effects and the Public Safety Tradeoff” (cit. on p. 74).

Rose, Evan K. (2021). “Who Gets a Second Chance? Effectiveness and Equity in Supervision of Criminal Offenders”. *Quarterly Journal of Economics* 136.2, pp. 1199–1253 (cit. on p. 75).

Sous-direction de la statistique et des études (2023). *Le Code de la justice pénale des mineurs, un premier bilan statistique à 15 mois*. Infostat Justice 194. Paris: Ministère de la Justice (SDSE) (cit. on p. 121).

Stevenson, Megan (2017). “Breaking Bad: Mechanisms of Social Influence and the Path to Criminality in Juvenile Jails”. *Review of Economics and Statistics* 99.5, pp. 824–838 (cit. on p. 74).

Stevenson, Megan T. and Jennifer L. Doleac (2024). “Algorithmic Risk Assessment in the Hands of Humans”. *American Economic Journal: Economic Policy* 16.4, pp. 382–414 (cit. on p. 75).

Tobón, Santiago (2022). “Do Better Prisons Reduce Recidivism? Evidence from a Prison Construction Program”. *Review of Economics and Statistics* 104.6, pp. 1256–1272 (cit. on p. 74).

Totarelli, Maddalena (2024). "Prison Rehabilitation Programs, Recidivism, and Labor Market Outcomes" (cit. on p. 74).

United Nations Children's Fund (2021a). *Achieving Justice for Children: A Review of Innovative Data Initiatives around the World*. Tech. rep. New York: UNICEF (cit. on p. 73).

— (2021b). *Estimating the Number of Children Deprived of Liberty in the Administration of Justice*. Tech. rep. New York: UNICEF (cit. on p. 73).

— (2025). *Children Deprived of Liberty in the Administration of Justice: Global and Regional Estimates and Trends*. Tech. rep. New York: UNICEF (cit. on p. 73).

Vuattoux, Arthur (2016). "Les centres éducatifs fermés pour les adolescents sont-ils une alternative à la prison?" *Mouvements* 88.4, pp. 117–123 (cit. on pp. 78, 80).

Youth Justice Board for England and Wales (2024). *Case Management Guidance: How to Use Reports* (cit. on p. 73).

Appendix

A The French Juvenile Justice System

A.1 Judicial Process and the Severity Ladder of Sanctions

Juvenile justice in France is administered by the *Protection Judiciaire de la Jeunesse* (PJJ), a directorate of the Ministry of Justice that operates public-sector facilities directly and accredits private associative facilities. From 1945 until September 2021 the system was governed by the *Ordonnance* of 2 February 1945, replaced by the *Code de la justice pénale des mineurs* (CJPM) on 30 September 2021. The two regimes share three founding principles: (i) educational primacy over punishment, (ii) specialised juvenile courts and prosecutors, and (iii) attenuated criminal responsibility scaled to age (Mucchielli, 2005). The analytical window of this paper, 2012-2022, is overwhelmingly pre-CJPM. The new code introduced strict statutory deadlines and a formal probationary period between guilt determination and sentencing but left the catalogue of judicial measures and the institutional landscape of placement options unchanged (Sous-direction de la statistique et des études, 2023; Cour des comptes, 2023).

The institutional pathway from offence to disposition unfolds in three stages described by Figure A1: a prosecutorial decision (drop, alternative to prosecution, or referral to a juvenile judge), a pre-trial phase in which the judge may order a socio-educational assessment and choose between community-based supervision, residential pre-trial placement or pre-trial detention, and a disposition hearing that closes the pre-trial phase by selecting an educational measure, an educational sanction, a post-judgment placement attached to a coercive judicial vehicle, or a criminal sentence. Detention enters the pathway through three doors: directly as a pre-trial measure, after the disposition through a firm custodial sentence, or as a downstream consequence of breaching a placement obligation under one of the coercive vehicles described below.

Across this pathway, the court selects from a hierarchy of increasing severity measures simplified in Figure A2. The least intrusive responses are alternatives to prosecution and community-based educational measures (warnings, reparation, supervised release, judicial educational measures). Residential placement comes next, escalating from individualised housing units and family hosts to collective foyers, reinforced educational centres (CER) and closed educational centres (CEF). Detention sits at the top of the ladder, either in juvenile units of adult prisons (QM) or in purpose-built juvenile detention centres (EPM). Most minors are handled at the lower end of the ladder: residential placement of any kind concerns only 6.6% of minors under penal measures (Direction de la Protection Judiciaire de la Jeunesse, 2024), and in the post-CJPM 2022 cohort, half of the convictions resulted in an educational measure and 8% in firm imprisonment (Sous-direction de la statistique et des études, 2023). Table A1 in Appendix A provides a complete summary of the measures referenced throughout the paper, with their target population, minimum age, and operational setting.

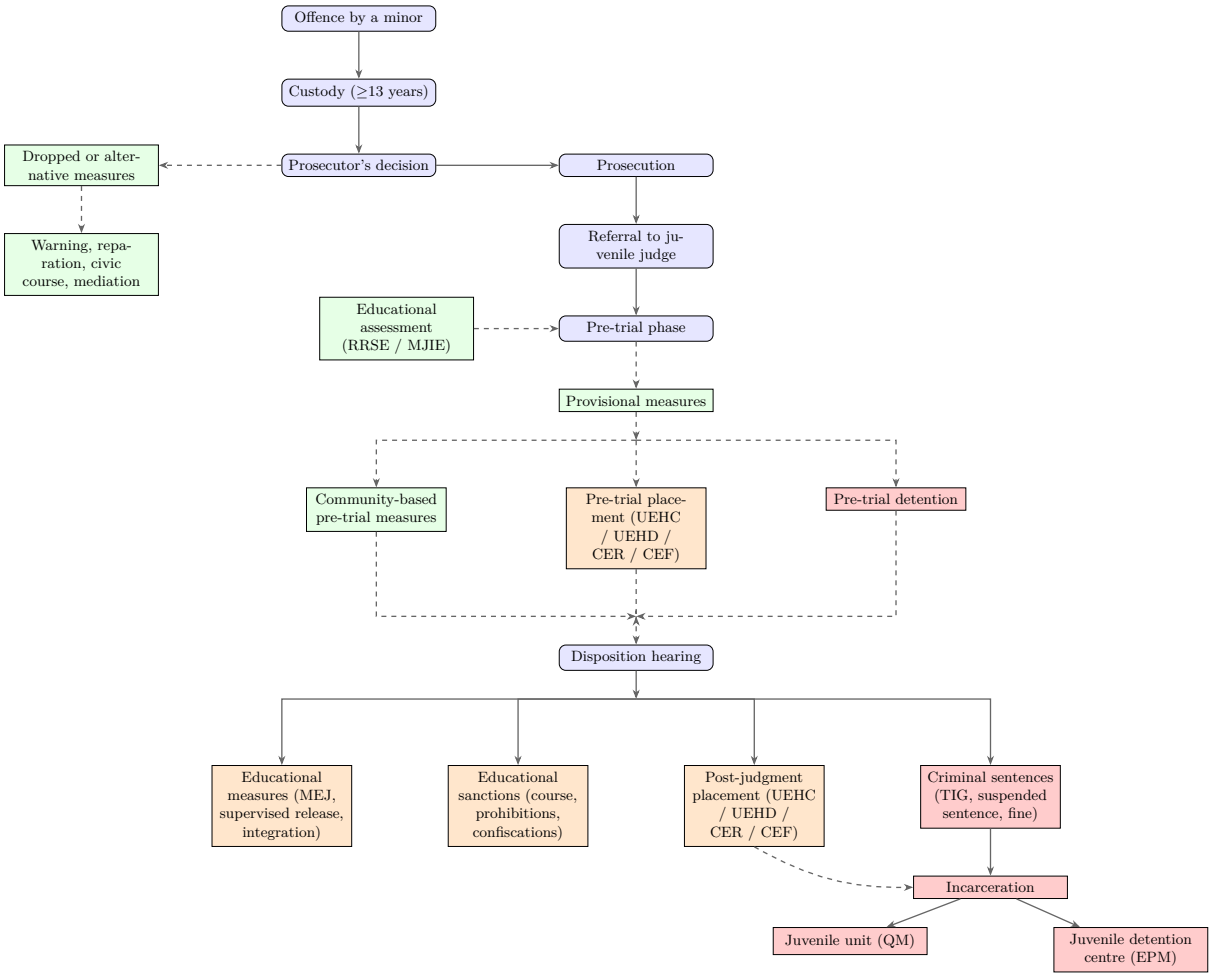

Figure A1: Simplified representation of the French juvenile penal pathway

Notes: Colours encode the role of each node. Blue marks judicial decision points (offence registration, prosecutor's decision, judge, hearings). Green marks educational measures and investigative tools (alternatives to prosecution, socio-educational assessment, community-based provisional measures). Orange marks post-decision placement and educational outcomes. Red marks custodial sentences and detention, including pre-trial detention. Solid arrows mark systematic transitions; dashed arrows mark optional or conditional ones. The diagram is a stylised summary of the pre-CJPM regime, and many minors are not routed through every step. During the pre-trial phase, the judge may order a socio-educational assessment of the minor and choose between three provisional alternatives, namely community-based supervision, a pre-trial residential placement, or pre-trial detention; the pre-trial phase may also end without any provisional measure and lead directly to the disposition hearing. At the disposition hearing, an educational sanction or educational measure can be ordered without any placement obligation, and post-judgment placement is itself contingent on the disposition attaching such an obligation to a coercive judicial vehicle. The dashed arrow from post-judgment placement to incarceration captures the breach mechanism described for closed educational centres.

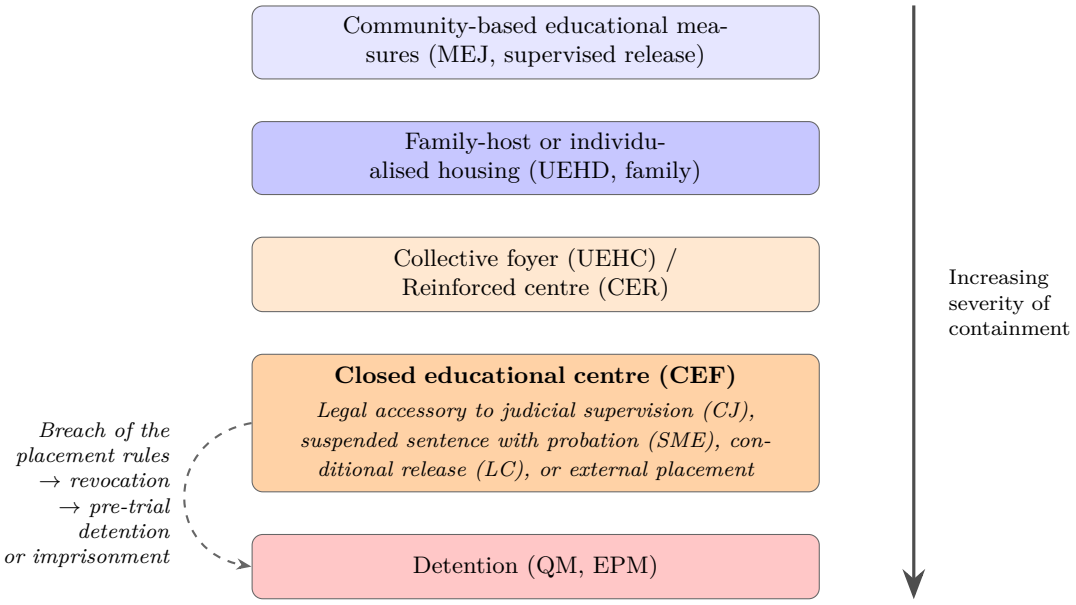

Figure A2: Severity ladder of placement decisions and the CEF as accessory to a coercive vehicle.

Notes: The ladder ranks the residential responses available to a juvenile court by intensity of containment, from community-based educational measures at the top to detention at the bottom. The CEF occupies the rung between collective foyers (UEHC) and reinforced educational centres (CER) on one side and detention on the other. The two dashed boxes capture the load-bearing legal facts of the dispositif: the CEF can only be ordered as an accessory to a coercive judicial vehicle (right-hand box), and a breach of its rules triggers a revocation of that vehicle that can convert the placement into pre-trial detention or imprisonment (left-hand box).

A.2 Judicial Measures Available to a Juvenile Court

Table A1: Main judicial measures in the French juvenile justice system

Measure (acronym)	**Target population**	**Min age**	**Purpose / setting**
Investigation			
Socio-Educational Inquiry Report (RRSE)	Any minor presented to a magistrate	–	Rapid socio-educational diagnostic produced by the duty PJJ educator at court before any pre-trial decision
Educational Judicial Investigation (MJIE)	Any minor under judicial proceedings	–	In-depth assessment of the minor's situation by a PJJ team over several months
Alternatives to prosecution			
Warning (*rappel à la loi*)	Simple facts, first-time offending	–	Formal reminder of the law without trial
Penal mediation (*médiation pénale*)	Inter-personal disputes	–	Restorative encounter between victim and offender
Civic course (*stage de citoyenneté*)	Emerging delinquency	10	Short educational programme on citizenship and the law
Reparation measure (REP)	Minor recognising responsibility	10	Concrete restorative action of a few hours to a few days, alone or in group
Community-based educational measures			
Judicial Educational Measure (MEJ)	Minor under PJJ supervision out of placement	–	Individualised educational follow-up by a PJJ educator
Educational measure pending judgment (*mise à l'épreuve éducative*)	Minor on the penal track under the CJPM regime	–	Structured observation between guilt determination and sentencing
Article 16bis educational measure (16B)	Minor in light delinquency	–	Short-format educational measure introduced by article 16bis of the 1945 ordinance
Daytime educational activity (MEAJ)	Minor under formal proceedings	–	Day-time programme of structured activities at a PJJ centre
Coercive judicial vehicles			
Judicial supervision (CJ)	Minor on the penal track, pre-trial or post-judgment	–	Court-mandated obligations and prohibitions; breach can trigger pre-trial detention
Suspended sentence with probation (SME)	Minor sentenced post-judgment	–	Conditional sentence with probation; breach can trigger sentence enforcement

Table A1: Main judicial measures in the French juvenile justice system *(continued)*

Measure (acronym)	**Target population**	**Min age**	**Purpose / setting**
Pre-trial supervised release (LSP)	Minor placed on conditional pre-trial release	–	Court-mandated obligations during the pre-trial phase
Supervised release (LS)	Minor under post-judgment community supervision	–	Court-mandated obligations after the disposition hearing
Sentence adjustment (AMP)	Sentenced minor near the end of incarceration	–	Early or modified release with continued supervision
Educational sanctions (SAE)			
Civic stage, prohibitions, confiscations	Targeted delinquency	10	Educational consequence of an offence; non-custodial
Residential placement (PJJ-operated)			
Individualised housing unit (UEHD)	Minor in social rupture, low-intensity supervision	–	Family-host or individual apartment; lower-intensity educational supervision
Collective foyer (UEHC)	Minor in placement, group setting	–	Collective foyer with 8-12 minors; round-the-clock educational supervision; daily cost approx. 323 €
Reinforced educational centre (CER)	Recidivist or rupture-prone minor	–	Three- to five-month residential rupture stay; up to 8 minors per facility; daily cost approx. 648 €
Closed educational centre (CEF)	Minor under judicial supervision or suspended sentence	13	Six-month placement, accessory to a coercive vehicle, breach triggers detention; 12 places per facility; daily cost approx. 705 € (public)
Criminal sentences (non-custodial)			
Community service order (TIG)	Older minor sentenced post-judgment	13	Unpaid work in a public-interest organisation
Suspended prison sentence	Minor sentenced post-judgment	13	Sentence not served unless reoffending occurs
Fine	Minor sentenced post-judgment	13	Pecuniary penalty
Detention			

Table A1: Main judicial measures in the French juvenile justice system *(continued)*

Measure (acronym)	**Target population**	**Min age**	**Purpose / setting**
Juvenile unit in adult prison (QM)	Minor sentenced or in pre-trial detention	13	Juvenile unit within adult prison; majority of detained minors; daily cost approx. 144 €
Juvenile detention centre (EPM)	Minor sentenced or in pre-trial detention	13	Purpose-built juvenile prison; up to 60 minors organised in small living units; six EPM operating in France

Notes: The table summarises the main judicial measures available to a juvenile court in France. Acronyms in parentheses are those used throughout the paper. Daily costs per minor are taken from Cour des comptes (2023) for 2021. The list is organised from the least to the most coercive response and is not exhaustive.

A.3 Geographic Roll-out of the Network

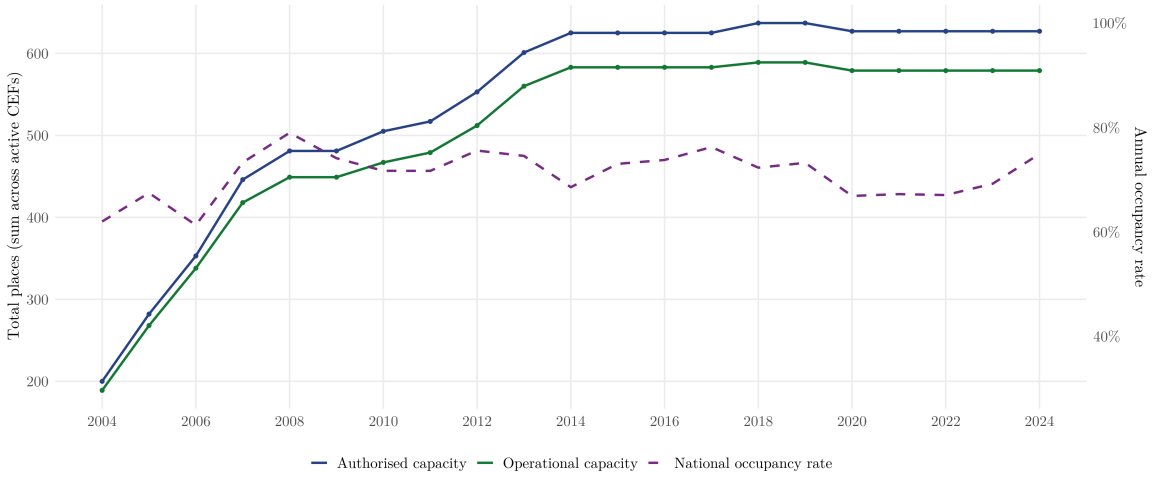

Figure A3: Network capacity and national occupancy of CEFs, 2004-2024.

Notes: The figure reports, on the left axis, the total authorised capacity (solid navy line) and the total operational capacity (solid green line) summed across all closed educational centres operating in each year, drawn from the DPJJ registry. On the right axis, the dashed violet line tracks the national mean of the centre-level annual occupancy rate, expressed as a share of authorised capacity. Source: DPJJ registry of accredited closed educational centres.

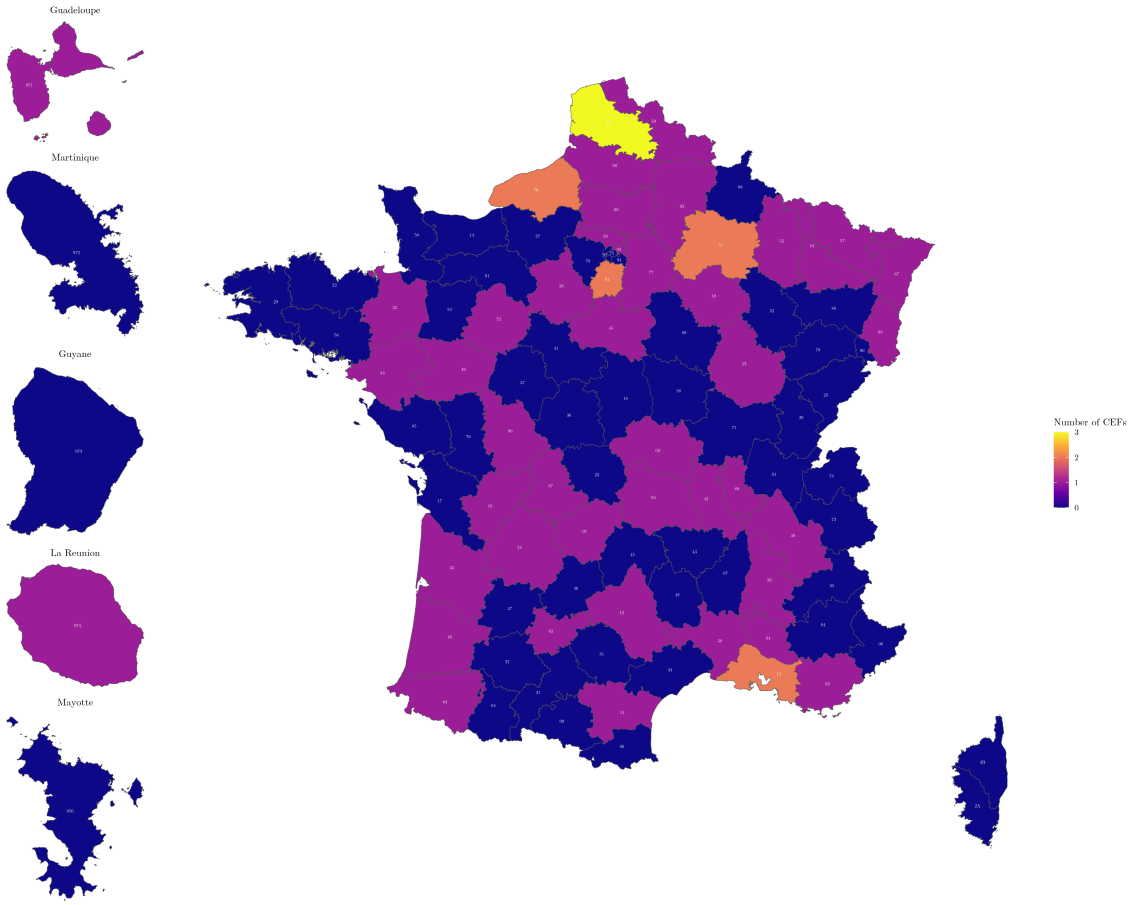

Figure A4: Geographic distribution of CEFs across French departments in 2022.

Notes: The map reports the number of closed educational centres operating in each French department at the end of 2022, drawn from the DPJJ registry. Metropolitan France appears on the right; the left-hand inset panels report the overseas departments where a CEF operates. The colour scale reads from no centre (light grey) to three or more centres (dark blue). Source: DPJJ registry of accredited closed educational centres.



Figure A5: Year-by-year geographic expansion of CEFs across French departments, 2003-2022.

Notes: The figure reports one map per year between 2003 and 2022, showing the number of closed educational centres operating in each French department at the end of the indicated year. Each panel uses the same colour scale, from no centre (light grey) to three or more centres (dark blue). The sequence visualises the three policy phases of the roll-out: a first wave between 2003 and 2010 in rural and peri-urban areas, an extension phase between 2011 and 2017 in denser urban areas, and a new-generation phase initiated in 2018. Source: DPJJ registry of accredited closed educational centres.

B Data Coverage and Measurement

B.1 Prescriber's Department Coverage (2004-2006)

The prescriber department identifier is partially missing in the early years of the GAME/IMAGE data. Before imputation, only 12% of spell records have an observed prescriber department in 2004, 26% in 2005, and 47% in 2006. From 2007 onward, coverage exceeds 99%. Because the department-level analyses assign each spell to its prescriber's department, missing department codes mechanically reduce the observed flow of measures and lower the computed rates in those early years.

Imputation procedure. I recover the prescriber department through a three-step procedure that exploits different sources of geographic information.

Step 1 Cross-temporal imputation (E1). For each minor whose prescriber department is missing in year t but observed in at least one other year, I assign the modal (most frequent) department from that minor's own records across all years. This step exploits the longitudinal structure of the data: most minors are followed by the same local PJJ unit across years. To validate this procedure, I mask the prescriber department for spells in 2007-2022 (where coverage is near-complete) and attempt to recover it using only the remaining records from other years. This cross-validation yields approximately 92% accuracy, confirming that department assignment is highly stable within individuals over time.

Step 2 Structure-based imputation (E2). For open-environment measures (investigations, community supervision, educational sanctions) where the prescriber department remains missing after Step 1, I use the department of the executing PJJ structure as a proxy. The structure department matches the prescriber department in 96-98% of cases for open-environment measures on the 2007-2022 subsample where both identifiers are observed (Table B2). This high match rate reflects the institutional logic: for non-residential interventions, the local PJJ service both prescribes and executes the measure. I do not apply this imputation to placement measures, where only 52% of spells have matching departments due to frequent inter-departmental transfers.

Step 3 Re-cross-imputation (E3). After Step 2 has added newly observed department information for open-environment spells, I rerun the cross-temporal imputation (Step 1) to propagate this information to remaining missing records - including placement spells that could not be imputed directly in Step 2 but whose minor now has additional department observations from open-environment measures.

Results. Table B1 reports the coverage rate at each imputation step for 2004-2010, and Table B2 reports the structure-prescriber match rates by measure type on the 2007-2022 validation subsample. Figure B1 plots the coverage trajectory across imputation levels. Step 1 alone raises

coverage from 12% to over 70% in 2004 and from 26% to over 80% in 2005, and the cumulative effect of the three steps raises coverage above 99% in all years.

Table B1: Coverage of prescriber department by imputation level

Year	N spells	E0 (%)	E0+E1 (%)	E0+E1+E2 (%)	All (%)
2004	103800	12.1	38.7	97.6	99.2
2005	123806	25.6	62.7	98.8	99.5
2006	133928	47.0	82.0	99.5	99.7
2007	142849	92.5	98.1	100.0	100.0
2008	145326	99.7	99.8	100.0	100.0
2009	158158	99.8	99.9	100.0	100.0
2010	176898	99.7	99.9	100.0	100.0

Notes: E0: original prescriber department from raw data. E1: cross-temporal imputation (modal department of same minor across years, 92

Table B2: Prescriber department vs structure department match rates by measure family

Measure family	N obs (2007-2022)	Match rate (%)	Used in E2
INV	2,163,429	97.9	Yes
SMO	1,778,125	96.4	Yes
PLA	191,703	50.0	No (51.9%)
AMP	65,013	97.1	Yes
SAE	47,113	98.0	Yes
MEAJ	1,482	98.0	Yes
ALT	983	99.5	Yes
CPE	667	98.8	Yes

Notes: Match rate between the prescriber department (judicial decision origin) and the structure department (facility location) on 2007-2022 data where both are observed. Open measures (INV, SMO, SAE, AMP, etc.) have 96-98

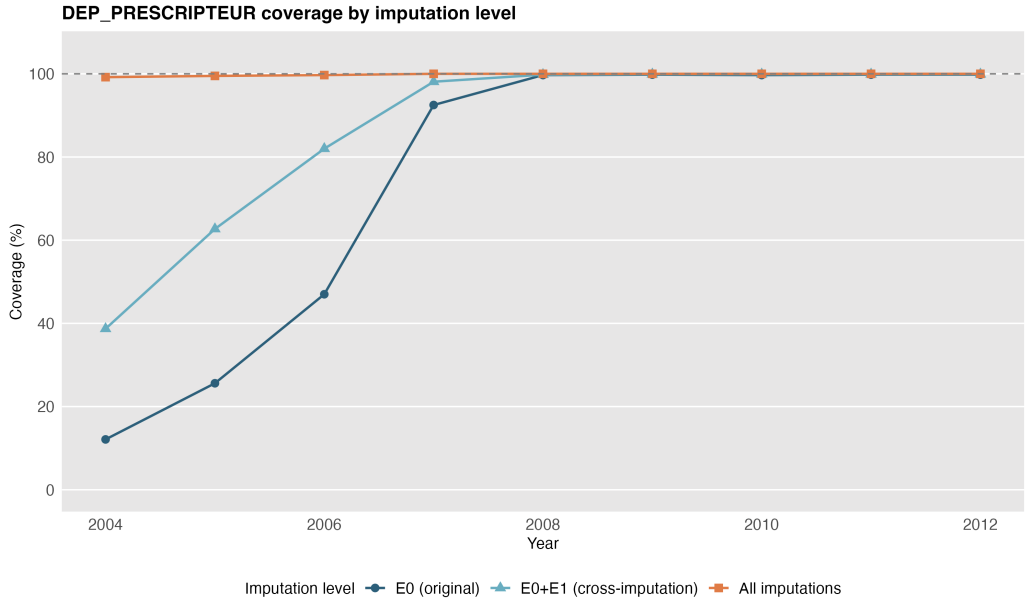

Figure B1: Prescriber department coverage by imputation level, 2004-2012.
Notes: This figure plots the share of judicial measure spells with a non-missing prescriber department for three imputation levels: E0 (original data), E0+E1 (after cross-temporal imputation), and all imputations combined (E0+E1+E2+E3). From 2007 onward, original coverage exceeds 95%. Source: PJJ administrative data (GAME/IMAGE), 2004-2012.

B.2 Detention: Data Availability (2010-2012)

Detention records in the GAME/IMAGE system were first extracted in 2012 and pulled back retroactively to approximately 2010. Table B3 reports the number of detention spells, the number of departments with at least one spell, and the mean detention rate for each year from 2004 to 2015, and Figure B2 displays the same counts graphically. Before 2008, the data contain fewer than 50 detention records per year - effectively zero. The ramp-up is steep: 341 records in 2009, 2,737 in 2010, 3,365 in 2011, and 3,757 in 2012, after which the series stabilises.

Table B3: Detention data coverage by year

Year	N spells	% of max	N depts (spells)	N depts (panel)	Mean detention rate
2005	1	0.0	1	0	NaN
2006	2	0.0	1	0	NaN
2007	3	0.0	1	0	NaN
2008	47	0.7	2	0	NaN
2009	341	5.0	6	0	NaN
2010	2737	39.8	50	101	0.00164
2011	3365	48.9	91	101	0.01289
2012	3757	54.6	98	101	0.02887
2013	3786	55.0	100	101	0.03000
2014	3540	51.4	97	101	0.02961
2015	3504	50.9	99	101	0.02922

Notes: This table reports detention data coverage from 2004 to 2015. N spells is the number of detention spells starting in each year from the raw GAME/IMAGE detention data.

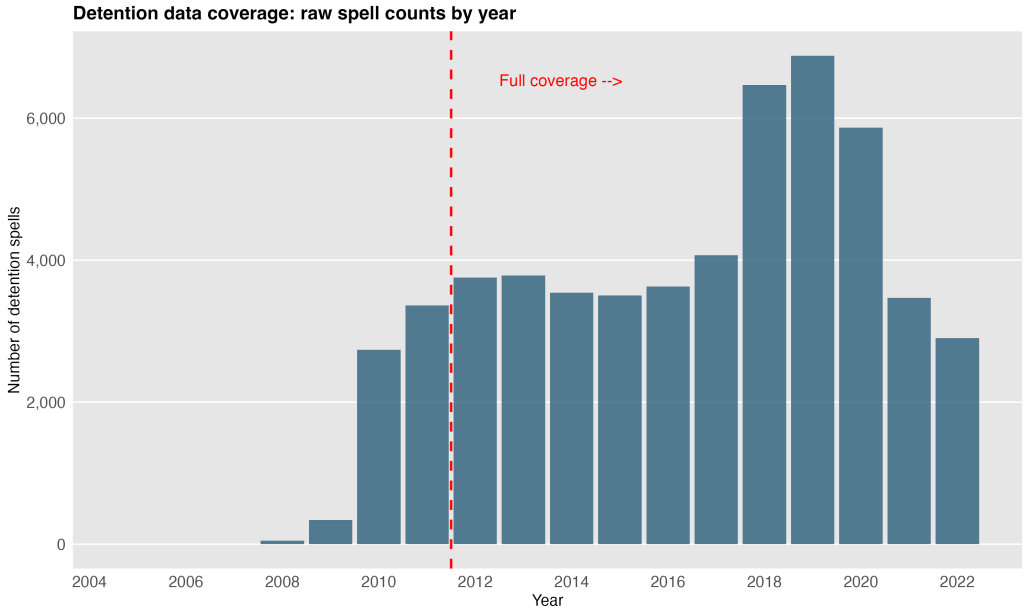

Figure B2: Detention data coverage: raw spell counts by year.
Notes: This figure displays the number of detention spells recorded per year in the GAME/IMAGE administrative data. The dashed red line marks 2012, the first year of complete coverage. Records before 2010 are negligible (fewer than 50 per year) and reflect retroactive data entry. Source: PJJ administrative data (GAME/IMAGE), 2004-2022.

This coverage limitation directly constrains the empirical strategy of the paper. Any outcome that loads on detention before 2012 is mechanically attenuated for the early treated cohorts, which is why the analytical window of the paper is restricted to 2012-2022 throughout (Section 3). To extend the analysis backwards and forwards, I am currently pursuing administrative access to the individual-level detention records held by the Direction de l'Administration Pénitentiaire (DAP). Once available, these records would close the pre-2010 coverage gap and would also allow me to follow the minors of the analytical sample after their eighteenth birthday into early adulthood, an outcome window that the DPJJ system stops recording at the age of majority.

Table B4: Duration of the indexed spell and age at its end, by arm and control type

	Median spell duration (days)	Mean spell duration (days)	Mean age at spell end	Share with unresolved end date
Eligible controls	312	393	17.1	0.000
Placed in a CEF	159	137	16.5	0.000
Other community supervision	218	279	17.0	0.000
Pre-trial supervised release	434	502	16.8	0.000
Supervised release (LS)	397	480	17.0	0.000
Residential placement	92	135	16.6	0.000
Sanctions and adjustments	153	177	16.2	0.000
Suspended sentence (SME)	620	593	18.4	0.000
Judicial supervision only	498	591	17.8	0.000
Article 16bis measure	563	585	17.9	0.000
Detention	60	96	17.1	0.000
Reinforced centre (CER)	100	99	16.4	0.000

Notes: This table reports the duration of the indexed spell and the minor's age when it ends, in the weighted common-support sample (control rows are unweighted descriptive statistics by control type). Durations are computed from the decision date to the recorded end of the spell; spells with an unresolved end date are reported as a share. The asymmetry in durations across arms motivates the decision-anchored outcomes: post-spell windows open at systematically different points of the trajectory on the two arms. Source: SSER data [GAME/IMAGES, 2012-2022].

B.3 Variable Definitions

Spell construction. Rows are collapsed into spells in two parallel pipelines. For measures, I collapse rows that describe administrative re-prescriptions of an already-running intervention rather than fresh judicial decisions. For detention, I merge consecutive rows of the same incarceration into one episode tagged by whether it is provisional only, conviction only, or provisional followed by conviction.

Structural exclusions. Three families of rows are removed from the raw output of the cleaning pipeline before any design choice applies. Civil-law procedures (codes 375 and PJM) are dropped because they correspond to child-protection rather than penal proceedings and follow a separate institutional logic. Investigation measures are dropped because they are diagnostic exercises producing a written report on the minor rather than interventions, and are systematically prescribed alongside other measures rather than as alternatives to a placement. Administrative wrappers, the supervision-style spells that double-count a placement when the legal vehicle and

the substantive intervention generate two rows, are dropped using a ±7-day proximity rule between the supervision row and the placement or detention row of the same minor; standalone instances of the same supervision measures, in which the minor truly received community supervision and not a placement, are kept.

Pre-spell controls. I construct twenty-one pre-spell covariates organised in four blocks: (i) demographics (sex, age at the indexed spell, age at the very first observed judicial measure); (ii) aggregate judicial activity prior to the indexed spell (count of all prior measures, count of those on the penal track); (iii) prior measures decomposed by type (residential placements - with separate counts for foyer, reinforced educational centre and all placements pooled - prior detentions, six community-supervision sub-types, prior investigations, and three other educational measure types); and (iv) the legal context in force at the time of the indexed spell (active judicial supervision order and active suspended sentence with probation, both with a seven-day buffer to ensure that an order prescribed jointly with the indexed spell does not contaminate the indicator). Each of these measure types is described in Table A1 in Appendix A. All prior-history counts are constructed by ordering the minor's full spell history chronologically and computing, for each indexed spell, the cumulative count of measures of the relevant type that started strictly before. Because the detention sub-base has very thin coverage before 2010, the prior-detention count is mechanically attenuated for early cohorts of the analytical window; I therefore add a twenty-second covariate, an indicator for a first observed spell predating 2010, which separates minors whose detention history is truncated by the coverage of the source from minors whose entire prior history falls under the modern one. It enters the balancing set on the same footing as the other twenty-one, so that placed and comparison minors are matched on the completeness of their observed history as well as on its content.

Local availability of a centre. I record an indicator equal to one if at least one closed educational centre was operating in the prescriber's department in the year of the indexed spell, and zero otherwise. The prescriber department is taken from the spell record; for the small share of detention spells in which the prescriber field is missing, I impute it from the trial court's department, which is recorded for the same spell.

A window is defined only when it closes before the end of the data and before the minor's eighteenth birthday, so the longer horizon of each source imposes the stricter requirement and drops the minors placed closest to majority. Sample sizes fall accordingly from one horizon to the next in every table. The binary outcomes cover three families, each measured on both horizons: a new detention episode, reconstructed by merging the provisional and the conviction segments of a single incarceration as described above; a new placement in a closed educational centre, which captures whether the placement returns the minor to the same modality; and a new criminal court referral, which proxies renewed penal involvement rather than a conviction.

B.4 Baseline Characteristics of the Analytical Cohorts

Table B5 reports the full set of baseline characteristics of the three cohorts discussed in Section 3.3: minors at their first observed placement of any kind, at their first closed educational centre placement, and at their first detention episode, all indexed between 2012 and 2022.

Table B5: Baseline characteristics of minors at first placement, first closed educational centre, and first detention

	First placement			First closed educational centre			First detention		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
Demographics									
Male	44,082	0.920	(0.272)	13,921	0.943	(0.231)	27,732	0.948	(0.221)
Foreign-born	37,961	0.200	(0.400)	12,086	0.167	(0.373)	26,338	0.289	(0.453)
Age at current event	44,082	16.297	(1.111)	13,921	16.227	(1.059)	27,732	16.810	(0.949)
Age at first judicial contact	44,082	15.351	(1.319)	13,921	14.982	(1.287)	27,732	15.458	(1.388)
Judicial history									
Number of prior judicial measures	44,082	4.047	(3.792)	13,921	7.098	(5.839)	27,732	7.047	(6.291)
Number of prior criminal court referrals	44,082	0.514	(1.028)	13,921	1.027	(1.565)	27,732	0.977	(1.546)
Placement									
Had a prior placement of any kind	44,082	0.000	(0.000)	13,921	0.454	(0.498)	27,732	0.383	(0.486)
... number of prior placements	0	–		6,314	1.903	(1.454)	10,619	2.224	(1.716)
Had a prior closed educational centre placement	44,082	0.000	(0.000)	13,921	0.000	(0.000)	27,732	0.178	(0.382)
... number of prior closed educational centre placements	0	–		0	–		4,928	1.246	(0.560)
Had a prior reinforced educational centre placement	44,082	0.000	(0.000)	13,921	0.146	(0.353)	27,732	0.095	(0.294)
... number of prior reinforced educational centre placements	0	–		2,031	1.131	(0.438)	2,648	1.156	(0.492)
Had a prior foyer or individualised housing placement	44,082	0.000	(0.000)	13,921	0.398	(0.489)	27,732	0.286	(0.452)
... number of prior foyer or individualised housing placements	0	–		5,539	1.755	(1.325)	7,922	1.817	(1.415)
Detention									
Had a prior detention	44,082	0.107	(0.310)	13,921	0.204	(0.403)	27,732	0.000	(0.000)
... number of prior detentions	4,738	1.204	(0.541)	2,839	1.280	(0.619)	0	–	
Community supervision									
Had a prior community supervision measure	44,082	0.656	(0.475)	13,921	0.757	(0.429)	27,732	0.704	(0.457)
... number of prior judicial supervision orders	28,902	0.581	(0.773)	10,545	1.119	(0.958)	19,512	1.174	(1.058)
... number of prior suspended sentences	28,902	0.111	(0.366)	10,545	0.223	(0.519)	19,512	0.339	(0.630)
... number of prior pre-trial supervised releases	28,902	0.758	(0.696)	10,545	0.806	(0.775)	19,512	0.818	(0.796)
... number of prior supervised releases	28,902	0.156	(0.414)	10,545	0.187	(0.453)	19,512	0.203	(0.482)
... number of prior reparation measures	28,902	0.694	(0.922)	10,545	0.789	(1.014)	19,512	0.746	(1.014)
... number of prior article 16bis educational measures	28,902	0.222	(0.479)	10,545	0.278	(0.541)	19,512	0.312	(0.576)
Investigation									
Had a prior investigation	44,082	0.738	(0.440)	13,921	0.795	(0.404)	27,732	0.748	(0.434)
... number of prior investigations	32,523	2.831	(2.101)	11,061	3.941	(2.790)	20,732	3.887	(2.999)
Other measures									
Had a prior sentence adjustment	44,082	0.028	(0.164)	13,921	0.047	(0.212)	27,732	0.050	(0.218)
... number of prior sentence adjustments	1,213	1.113	(0.358)	659	1.149	(0.400)	1,386	1.151	(0.418)
Had a prior educational sanction	44,082	0.021	(0.144)	13,921	0.032	(0.175)	27,732	0.032	(0.176)
... number of prior educational sanctions	931	1.125	(0.404)	442	1.152	(0.464)	891	1.147	(0.444)
Had a prior daytime educational activity	44,082	0.002	(0.044)	13,921	0.002	(0.049)	27,732	0.003	(0.053)
... number of prior daytime educational activities	87	1.080	(0.313)	34	1.059	(0.239)	79	1.051	(0.221)
Current spell type									
Duration of current event (days)	44,082	139.108	(139.442)	13,921	137.334	(88.363)	27,732	93.815	(139.852)
Active judicial supervision at current event	44,082	0.229	(0.420)	13,921	0.450	(0.497)	27,732	0.415	(0.493)
Active suspended sentence at current event	44,082	0.057	(0.232)	13,921	0.127	(0.333)	27,732	0.175	(0.380)
Minors	44,082			13,921			27,732		

Notes: This table reports the characteristics of three cohorts of minors, each defined by the first event of the indicated type recorded between 2012 and 2022. Minors whose first event of that type falls outside this window are excluded from the corresponding cohort. The three cohorts overlap: a minor whose first placement is in a closed educational centre appears in both the first placement and the first closed educational centre columns. Means are proportions for binary variables and raw means otherwise, with the standard deviation in parentheses. Rows starting with an ellipsis are conditional means computed on the minors who had at least one prior measure of the corresponding type, hence the smaller counts. By construction the first placement cohort has no prior placement, the first closed educational centre cohort no prior placement in such a centre, and the first detention cohort no prior detention. Foyer and individualised housing covers the non-specialised residential units, as distinct from the reinforced and closed centres counted separately. Article 16bis educational measures are short-format educational measures introduced by article 16bis of the 1945 ordinance on juvenile justice. Sentence adjustments modify how a custodial sentence is served. Educational sanctions are penalties available from age ten. Daytime educational activity is a structured weekday programme created in 2007. Country of birth is recorded for a subset of minors only, and the count column reports how many carry a recorded value. Source: SSER data [GAME/IMAGES, 2012-2022].

B.5 Selection into the Report-Linked Sample

The recommendation-based designs of Section 5.2 are estimated on the measures whose social-educational report is linked to them. Table B6 describes how that subset is selected, and Table B7 asks whether the linkage of the report predicts what follows the decision to place.

Table B6: Measures with and without a linked social-educational report (PARCOURS, 2021-2024)

	Linked report	No linked report	Std. difference
Panel A: All eligible measures			
Placed in a closed educational centre	0.018	0.009	0.076
Age at the measure	16.643	15.507	0.508
Male	0.942	0.796	0.445
Prior judicial episodes	6.982	3.441	0.518
Prior measures on the penal track	6.719	3.236	0.526
Prior residential placements	0.526	0.303	0.194
Prior detention episodes	0.244	0.114	0.181
Prior closed centre placements	0.136	0.073	0.141
Measures	80,163	250,812	
Panel B: Closed educational centre placements only			
Age at the measure	16.283	16.280	0.003
Male	0.953	0.949	0.022
Prior judicial episodes	12.388	9.925	0.290
Prior measures on the penal track	11.909	9.481	0.293
Prior residential placements	1.427	1.371	0.032
Prior detention episodes	0.520	0.449	0.066
Prior closed centre placements	0.417	0.435	-0.022
Measures	1,460	2,341	

Notes: This table compares, within the population eligible for the recommendation-based analysis, measures whose attached social-educational report is linked to them with measures carrying no linked report. Panel A covers all eligible penal measures indexed from 2021; Panel B restricts the comparison to measures executed in a closed educational centre. Each row reports the mean of the indicated characteristic in the two groups; the last column is the difference in means over the pooled standard deviation. A linear probability model of the linkage indicator on the characteristics of the minor alone yields an adjusted R squared of 0.085, against 0.206 for court-department by year fixed effects alone; adding the characteristics on top of those fixed effects raises it to 0.270, and among placements the same addition moves it from 0.229 to 0.255. A measure executed in a closed educational centre is 14.4 percentage points more likely to carry a linked report than another measure; the gap falls to 6.2 points (standard error 0.9) once the fixed effects and the characteristics of the minor are absorbed, so linkage remains more frequent for placements than the rest of the flow. Prior-history counts are cumulative counts of measures started strictly before the indexed one. Source: SSER data [PARCOURS, 2021-2024].

Table B7: Subsequent judicial events of closed centre placements, with and without a linked report

Outcome	Coefficient	Standard error	Mean without a linked report	N
Detention episode, 12 months	0.019	(0.022)	0.452	1,867
Detention episode, 18 months	0.011	(0.037)	0.525	1,033
New court referral, 12 months	0.062***	(0.019)	0.857	1,867
New court referral, 18 months	0.032	(0.020)	0.900	1,033

Notes: This table restricts the sample to measures executed in a closed educational centre and asks whether the linkage of the social-educational report predicts what follows the decision to place. Each row is a separate linear probability model of the indicated outcome on an indicator for a linked report, the demographic and prior-history characteristics of the minor, and court-department by year fixed effects. Outcomes are anchored at the decision date over fixed windows and are defined only where the window closes before the end of the data and the minor's eighteenth birthday. The fourth column reports the mean of the outcome among placements without a linked report. A coefficient close to zero provides no evidence that linkage selects placements on the indicated subsequent outcome, conditional on the recorded trajectory and court-by-year fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, in parentheses, are clustered at the court-department level. Source: SSER data [PARCOURS, 2021-2024].

C Comparison Group, Balance and Sensitivity

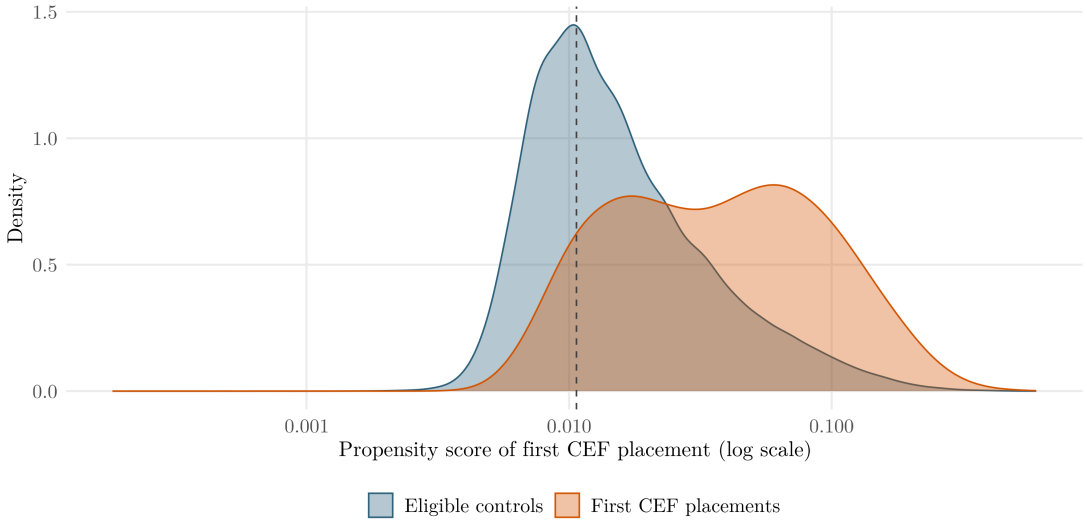

Figure C1: Distribution of the placement propensity score in the two arms.

Notes: The figure reports the density of the estimated propensity of receiving a first closed educational centre placement, on a logarithmic scale, separately for first closed centre placements and for the eligible untreated spells. The score is estimated by logit on the pre-spell covariates and year effects. The dashed vertical line marks the tenth percentile of the score among placed minors, which is the common-support threshold applied to both arms alike and which retains 90 percent of placements. Sample: spells recorded in the historical registers of the French juvenile justice directorate between 2012 and 2022.

Table C1: Covariate balance before and after entropy-balancing reweighting, common-support sample

		Before weighting			After weighting		
Covariate	Treated mean	Control mean	Std. diff.	p-value	Control mean	Std. diff.	p-value
Prior judicial episodes (any)	8.09	5.27	0.53	0.000	8.09	0.00	1.000
Prior criminal court referrals	1.14	0.72	0.28	0.000	1.14	0.00	1.000
Prior foyer placements	0.78	0.38	0.38	0.000	0.78	0.00	1.000
Prior reinforced centre placements	0.17	0.07	0.27	0.000	0.17	0.00	1.000
Prior judicial supervision orders	0.98	0.49	0.54	0.000	0.98	0.00	1.000
Prior suspended sentences with probation	0.19	0.11	0.16	0.000	0.19	0.00	1.000
Prior pre-trial supervised releases	0.69	0.52	0.24	0.000	0.69	0.00	1.000
Prior supervised releases	0.16	0.12	0.10	0.000	0.16	0.00	1.000
Prior reparation measures	0.66	0.52	0.15	0.000	0.66	0.00	1.000
Prior article 16bis measures	0.24	0.16	0.16	0.000	0.24	0.00	1.000
Prior sentence adjustments	0.05	0.04	0.06	0.000	0.05	0.00	1.000
Prior educational sanctions	0.04	0.03	0.04	0.003	0.04	0.00	1.000
Prior daytime educational activities	0.00	0.00	0.01	0.695	0.00	0.00	1.000
Prior investigation measures	3.53	2.42	0.41	0.000	3.53	0.00	1.000
Age at the indexed spell	16.11	16.06	0.05	0.058	16.11	0.00	1.000
Age at first judicial contact	14.74	14.96	-0.19	0.000	14.74	-0.00	1.000
Active judicial supervision order	0.54	0.26	0.56	0.000	0.54	0.00	1.000
Active suspended sentence	0.15	0.08	0.18	0.000	0.15	0.00	1.000
Male	0.95	0.94	0.05	0.003	0.95	0.00	1.000
Prior detention episodes	0.29	0.15	0.24	0.000	0.29	0.00	1.000
Prior placements (any type)	0.98	0.47	0.41	0.000	0.98	0.00	1.000
First observed spell before 2010	0.02	0.02	0.02	0.019	0.02	0.00	1.000

Notes: This table compares treated first CEF placements with all eligible controls on every covariate entering the balancing procedure, in the common-support sample of the main specification. For each covariate, the table reports the treated mean, the control mean, the standardised difference (mean difference divided by the treated standard deviation) and the p-value of the difference, before weighting and after entropy-balancing reweighting. P-values come from regressions of the covariate on the treatment indicator with standard errors clustered at the prescriber-department level, weighted in the after-weighting columns. Entropy balancing equalises means exactly, so post-weighting differences are zero up to numerical precision. Count covariates are capped at their 99th percentile. Source: SSER data [GAME/IMAGES, 2012-2022].

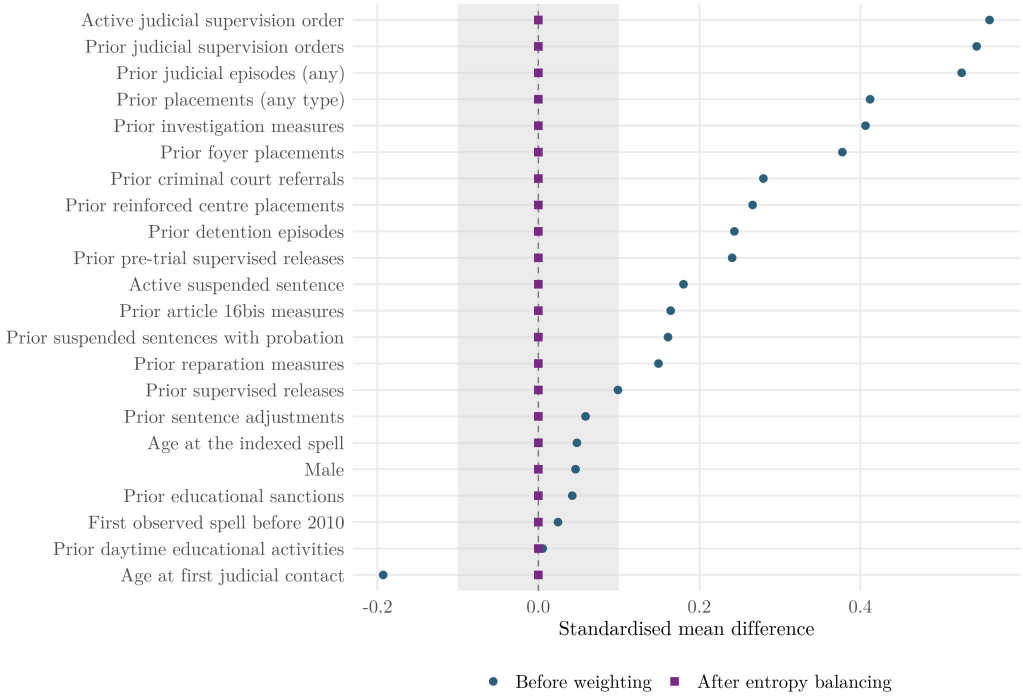

Figure C2: Standardised covariate differences before and after reweighting.

Notes: The figure reports, for each covariate entering the balancing set, the standardised mean difference between placed minors and untreated spells, before any weighting and after entropy-balancing reweighting. A standardised mean difference is the difference in means divided by the standard deviation among placed minors. The shaded band marks the conventional tolerance of one tenth of a standard deviation and the dashed line marks exact equality. Sample: spells recorded in the historical registers of the French juvenile justice directorate between 2012 and 2022, restricted to the region of common support of the propensity score.

Table C2: Composition of the comparison group on common support

Measure received by the comparison spell	Spells	Share
Community supervision (77.5%)		
Judicial supervision only	53,576	15.8%
Pre-trial supervised release	73,958	21.8%
Suspended sentence (SME)	19,971	5.9%
Supervised release (LS)	20,717	6.1%
Article 16bis measure	31,718	9.3%
Sanctions and adjustments	25,323	7.5%
Other community supervision	38,090	11.2%
Residential placement (15.6%)		
Residential placement	43,301	12.7%
Reinforced centre (CER)	9,731	2.9%
Detention (6.8%)		
Detention	23,237	6.8%
Total	339,622	100.0%

Notes: This table reports the measure actually received by the untreated spells of the common-support sample, that is, the population against which placements are compared in the pooled estimates. The sample covers 2012-2022 spells whose propensity score exceeds the tenth percentile of the treated score distribution. Shares are computed over all untreated spells on support. The distribution is the GAME counterpart of the destinations observed in PARCOURS among minors recommended for a closed educational centre and redirected elsewhere. Source: SSER data [GAME/IMAGES, 2012-2022].

Table C3: Selection-on-unobservables statistics for the main comparison

Outcome	Unadjusted coef.	Adjusted coef.	R^2 movement	δ for zero effect	Bias-adjusted coef. ($\delta = 1$)
Detention episode, 12 months	0.321	0.198	$0.043 \rightarrow 0.168$	4.0	0.149
Detention episode, 18 months	0.363	0.205	$0.050 \rightarrow 0.183$	3.1	0.140
New court referral, 12 months	0.317	0.155	$0.038 \rightarrow 0.153$	2.4	0.090
New court referral, 18 months	0.318	0.132	$0.045 \rightarrow 0.157$	1.7	0.054

Notes: This table applies the coefficient-stability statistics of Oster (2019) to the comparison of all eligible controls with placed minors on the common-support sample, using unweighted OLS with year and prescriber-department fixed effects (the statistics are derived for OLS; the entropy-balanced estimates are therefore not used here). The delta column reports how many times stronger than the observed covariates the selection on unobservables would need to be, proportionally, to drive the adjusted coefficient to zero, with the maximum attainable R-squared set to 1.3 times the adjusted one; the last column reports the bias-adjusted coefficient when unobservable selection is as strong as observable selection. Values of delta above one indicate estimates robust to proportional selection, following the interpretation convention of Oster (2019); the formulas use the proportional-selection approximation. Source: SSER data [GAME/IMAGES, 2012-2022].

Table C4: Strength an unobserved factor would need to overturn the estimate, report-linked sample and recommended population

Outcome	Adjusted coef.	δ for zero effect	Bias-adjusted coef. ($\delta = 1$)	Robustness value	Robustness value at the 5% level	Recommendation: placement	Recommendation: detention	Covariates: placement	Covariates: detention	Multiple of the recommendation
Panel A: report-linked sample										
Detention episode, 12 months	0.139	1.9	0.067	4.4%	3.3%	5.7%	0.3%	2.6%	21.1%	2.9
Detention episode, 18 months	0.087	0.9	-0.012	2.8%	1.3%	6.1%	0.6%	3.4%	24.4%	1.3
Panel B: recommended population										
Detention episode, 12 months	0.086	-13.8	0.092	7.1%	2.6%	–	–	6.3%	16.4%	–
Detention episode, 18 months	0.041	5.8	0.034	4.0%	0.0%	–	–	10.1%	22.8%	–

Notes: This table reports how strong selection on factors the data do not record would have to be to overturn the estimated effect of a first closed educational centre placement. Panel A uses every penal measure whose attached social-educational report is linked, where the recommendation varies and enters as a control. Panel B restricts to the measures whose report recommended a closed educational centre, which is the population the headline estimate is computed on; the recommendation is constant there by construction and leaves the conditioning set, so the columns that use it as a yardstick are empty. The adjusted coefficient comes from an unweighted linear regression of the outcome on the placement indicator, the seventeen pre-measure covariates, the recommendation indicators where they vary, and year and court-department fixed effects, with standard errors clustered at the court-department level. Three quantities follow, and they answer different questions. The first, due to Oster (2019), assumes that selection on unrecorded factors is proportional to selection on the recorded ones: the delta column gives the ratio at which the coefficient would reach zero, and the next column gives the coefficient when the two are equally strong, with the maximum attainable R-squared set to 1.3 times the adjusted one. A negative delta arises when the coefficient moves away from zero as the covariates are added, so that proportional selection would amplify the estimate rather than explain it away. The two following columns are both robustness values in the sense of Cinelli and Hazlett (2020), and they are not interchangeable. The first is the share of the variation left unexplained in both the placement and the outcome that a single unrecorded factor would have to account for, jointly, to bring the coefficient exactly to zero. The second is the share required only to push it to the edge of significance at the five per cent level, which is a lower bar and is zero whenever the estimate is already not significant at that level. The last three columns give benchmarks against which those shares can be read, since a share is otherwise uninterpretable on its own. They report how much of the same residual variation the educator's recommendation accounts for, on the placement margin and on the outcome margin, and how much the block of recorded covariates accounts for on each. Comparing a robustness value with one of these directly would be misleading, since a factor can be far more predictive of placement than of the outcome; the final column resolves this for the recommendation by reporting the multiple of it that brings the coefficient to zero when both margins are scaled together, a value above one meaning an unrecorded factor would have to be more predictive than the recommendation on both margins at once. Source: SSER data [PARCOURS, 2021-2024].

Table C5: Placebo balance on pre-determined variables and on outcomes measured before the decision

	Difference	SE	Counterfactual mean	Observations
Days since the latest prior judicial contact	-70.692***	(6.242)	-420.29	314,770
Duration of the previous spell (days)	-36.016***	(5.355)	212.98	314,770
Previous spell was a criminal court referral	-0.015**	(0.007)	0.20	314,770
Detention episode in the twelve months before the decision	0.053***	(0.007)	0.19	351,185
Court referral in the twelve months before the decision	0.090***	(0.009)	0.38	351,185

Notes: This table tests the exchangeability assumption on quantities the adjustment was not asked to equalise. The first three rows report weighted treated-control differences on pre-determined variables deliberately left out of the balancing set: if balancing on the covariates it does use were enough to align the two arms on everything that matters, these differences would be close to zero. The last two rows apply the same weighted comparison to the outcomes themselves, measured over the twelve months preceding the decision date and closing strictly before it. The placement cannot have caused what happened before it was ordered, so a non-zero difference there measures residual selection that the weights leave in place rather than an effect. All differences are estimated on the common-support sample with entropy-balancing weights, year and prescriber-department fixed effects, and standard errors clustered at the prescriber-department level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Source: SSER data [GAME/IMAGES, 2012-2022].

Table C6: Sensitivity of the detention estimates to the imbalanced pre-determined variables

	Baseline	SE	Adding the three variables	SE	Observations
Detention episode, 12 months (decision anchor)	0.208***	(0.011)	0.203***	(0.011)	220,350
Detention episode, 18 months (decision anchor)	0.220***	(0.011)	0.215***	(0.011)	163,014
Detention episode, 24 months (decision anchor)	0.222***	(0.013)	0.217***	(0.013)	112,671

Notes: This table re-estimates the main-panel detention effects on the sub-sample of spells preceded by at least one prior judicial contact, with the entropy-balancing weights of the main analysis, and compares the baseline specification with one adding the three imbalanced pre-determined variables of the placebo table (recency of the latest prior contact, duration and penal status of the previous spell) as regression controls. Stability of the coefficient indicates that the documented imbalance, while real, does not carry the estimate. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, in parentheses, are clustered at the prescriber-department level. Source: SSER data [GAME/IMAGES, 2012-2022].

Table C7: Where the adjustment comes from: raw, covariate-adjusted, inverse-probability weighted and entropy-balanced estimates

	Detention episode		New CEF placement		New court referral	
	12 months	18 months	12 months	18 months	12 months	18 months
Fixed effects only	0.321***	0.363***	0.201***	0.243***	0.317***	0.318***
	(0.012)	(0.012)	(0.012)	(0.013)	(0.009)	(0.010)
Adding covariates	0.198***	0.205***	0.133***	0.148***	0.155***	0.132***
	(0.010)	(0.010)	(0.011)	(0.012)	(0.008)	(0.009)
Inverse-probability weights only	0.242***	0.275***	0.138***	0.168***	0.205***	0.207***
	(0.011)	(0.011)	(0.011)	(0.013)	(0.009)	(0.010)
Inverse-probability weights and covariates	0.205***	0.214***	0.134***	0.155***	0.169***	0.153***
	(0.011)	(0.011)	(0.011)	(0.013)	(0.009)	(0.009)
Entropy weights only	0.242***	0.276***	0.143***	0.174***	0.209***	0.212***
	(0.010)	(0.011)	(0.011)	(0.013)	(0.009)	(0.010)
Entropy weights and covariates	0.204***	0.215***	0.130***	0.148***	0.166***	0.149***
	(0.010)	(0.011)	(0.011)	(0.013)	(0.008)	(0.009)

Notes: Each row is one specification estimated on the same common-support sample with all eligible controls: the raw contrast with year and prescriber-department fixed effects; the same with the twenty-two pre-spell covariates as linear controls and no reweighting; then two families of weights, each with and without the covariates. Inverse-probability weights are the classical propensity-score weights for the effect on the treated - unit weight for placed minors, the odds of the estimated score for controls, trimmed at the ninety-ninth percentile - which balance covariates in expectation. Entropy-balancing weights instead impose exact equality of the covariate means by construction, and their combination with the covariates is the doubly robust specification used throughout the paper. Agreement between the two weighting families indicates that the estimate does not depend on how the propensity score is used. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, in parentheses, are clustered at the prescriber-department level. Source: SSER data [GAME/IMAGES, 2012-2022].

Table C8: Common-support rules and resulting analytical samples

Rule	Cutoff	Spells kept	Treated kept	Share of treated kept	Treated share in sample
No support restriction	0.000	551,929	12,848	100.0%	2.3%
p10 of treated scores	0.011	351,185	11,563	90.0%	3.3%
p25 of treated scores	0.017	214,031	9,637	75.0%	4.5%
Median of treated scores	0.037	81,909	6,424	50.0%	7.8%
Top decile of the pool	0.050	55,193	5,184	40.3%	9.4%
Top 5 percent of the pool	0.076	27,597	3,205	24.9%	11.6%

Notes: This table reports, for five candidate common-support rules, the threshold value of the propensity score, the number of spells and of first CEF placements retained, and the share of treated spells in the resulting sample. The propensity score is estimated on the full 2012-2022 pool by logistic regression on twenty-one pre-spell covariates, a truncated-history cohort indicator and year fixed effects. The main specification uses the tenth percentile of the treated score distribution; the other rules are reported as sensitivity checks. Source: SSER data [GAME/IMAGES, 2012-2022].

Table C9: Decision-anchored effect within the recommended population, alternative fixed effects

	Detention episode		New CEF placement		New court referral	
	12 months	18 months	12 months	18 months	12 months	18 months
Panel A: Prescriber department and year						
Placed in a closed educational centre	0.080***	0.042	-0.056**	-0.060*	0.011	-0.024*
	(0.027)	(0.034)	(0.028)	(0.033)	(0.016)	(0.013)
Counterfactual mean	0.435	0.567	0.258	0.331	0.949	0.972
Observations	1,737	897	1,737	897	1,737	897
Treated observations	286	163	286	163	286	163
Adjusted R^2	0.230	0.325	0.198	0.274	0.120	0.214
Panel B: Court and year						
Placed in a closed educational centre	0.067**	0.026	-0.054*	-0.063*	0.015	-0.022
	(0.027)	(0.034)	(0.028)	(0.038)	(0.016)	(0.015)
Counterfactual mean	0.435	0.567	0.258	0.331	0.949	0.972
Observations	1,737	897	1,737	897	1,737	897
Treated observations	286	163	286	163	286	163
Adjusted R^2	0.236	0.376	0.195	0.265	0.129	0.218
Panel C: Court by year						
Placed in a closed educational centre	0.061**	0.008	-0.049	-0.053	0.026	-0.017
	(0.024)	(0.026)	(0.030)	(0.036)	(0.017)	(0.015)
Counterfactual mean	0.435	0.567	0.258	0.331	0.949	0.972
Observations	1,737	897	1,737	897	1,737	897
Treated observations	286	163	286	163	286	163
Adjusted R^2	0.320	0.435	0.270	0.309	0.203	0.371
Included in all regressions:						
Pre-spell judicial history and demographic controls	Y	Y	Y	Y	Y	Y

Notes: This table reports weighted least squares estimates of the average effect of a first closed educational centre placement on the treated. Each panel restricts the control group to the indicated type of alternative measure; unless stated otherwise the treated group is identical across panels and the entropy-balancing weights are re-estimated within each panel so that the restricted control pool matches the treated covariate profile (weighting method by panel: Prescriber department and year: Entropy; Court and year: Entropy; Court by year: Entropy). Each column corresponds to one post-spell outcome and horizon; outcomes at twelve months are defined only for spells whose full one-year post-spell window is observable before the data end and the minor's eighteenth birthday, and therefore over-represent minors placed young. The counterfactual mean is the weighted control mean of the outcome. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, in parentheses, are clustered at the prescriber-department level. Each panel re-estimates the same specification with a different set of absorbed effects, from the prescriber department to the deciding court and then to the court by year cell. Standard errors remain clustered at the prescriber-department level throughout. Source: SSER data [PARCOURS, 2021-2024].

Table C10: Effect of a first CEF placement, control decisions down-weighted by minor, twelve and eighteen months from the decision

	Detention episode		New CEF placement		New court referral	
	12 months	18 months	12 months	18 months	12 months	18 months
Panel A: All eligible controls						
Placed in a closed educational centre	0.232***	0.249***	0.106***	0.123***	0.239***	0.233***
	(0.011)	(0.011)	(0.011)	(0.013)	(0.009)	(0.009)
Counterfactual mean	0.185	0.204	0.168	0.195	0.397	0.448
Observations	255,021	193,050	255,021	193,050	255,021	193,050
Treated observations	8,381	6,196	8,381	6,196	8,381	6,196
Adjusted R^2	0.234	0.281	0.146	0.186	0.247	0.283
Panel B: Community supervision (all)						
Placed in a closed educational centre	0.223***	0.240***	0.129***	0.145***	0.254***	0.238***
	(0.011)	(0.011)	(0.012)	(0.014)	(0.009)	(0.009)
Counterfactual mean	0.192	0.214	0.147	0.179	0.385	0.452
Observations	185,759	142,784	185,759	142,784	185,759	142,784
Treated observations	8,381	6,196	8,381	6,196	8,381	6,196
Adjusted R^2	0.248	0.282	0.136	0.171	0.218	0.240
Panel C: Judicial supervision only						
Placed in a closed educational centre	0.162***	0.170***	0.080***	0.094***	0.119***	0.093***
	(0.011)	(0.011)	(0.012)	(0.014)	(0.011)	(0.012)
Counterfactual mean	0.258	0.305	0.196	0.242	0.522	0.620
Observations	47,398	34,338	47,398	34,338	47,398	34,338
Treated observations	8,381	6,196	8,381	6,196	8,381	6,196
Adjusted R^2	0.202	0.230	0.117	0.139	0.162	0.176
Panel D: Pre-trial supervised release						
Placed in a closed educational centre	0.255***	0.285***	0.167***	0.187***	0.201***	0.179***
	(0.019)	(0.015)	(0.017)	(0.020)	(0.014)	(0.014)
Counterfactual mean	0.178	0.179	0.106	0.133	0.449	0.521
Observations	78,793	68,057	78,793	68,057	78,793	68,057
Treated observations	8,381	6,196	8,381	6,196	8,381	6,196
Adjusted R^2	0.313	0.263	0.149	0.162	0.276	0.230
Panel E: Suspended sentence / supervised release						
Placed in a closed educational centre	0.220***	0.226***	0.154***	0.176***	0.263***	0.249***
	(0.012)	(0.013)	(0.012)	(0.015)	(0.011)	(0.011)
Counterfactual mean	0.199	0.236	0.125	0.158	0.399	0.475
Observations	38,079	27,625	38,079	27,625	38,079	27,625
Treated observations	8,381	6,196	8,381	6,196	8,381	6,196
Adjusted R^2	0.230	0.279	0.122	0.153	0.190	0.216
Panel F: Residential placement (excl. CER)						
Placed in a closed educational centre	0.147***	0.151***	0.032**	0.036**	0.085***	0.065***
	(0.010)	(0.011)	(0.013)	(0.014)	(0.011)	(0.012)
Counterfactual mean	0.285	0.336	0.259	0.319	0.583	0.675
Observations	38,470	28,214	38,470	28,214	38,470	28,214
Treated observations	8,381	6,196	8,381	6,196	8,381	6,196
Adjusted R^2	0.159	0.191	0.089	0.102	0.108	0.142
Panel G: Reinforced educational centre (CER)						
Placed in a closed educational centre	0.173***	0.157***	0.064***	0.050**	0.118***	0.101***
	(0.013)	(0.016)	(0.016)	(0.021)	(0.010)	(0.011)
Counterfactual mean	0.255	0.322	0.220	0.305	0.551	0.639
Observations	15,439	11,200	15,439	11,200	15,439	11,200
Treated observations	8,381	6,196	8,381	6,196	8,381	6,196
Adjusted R^2	0.169	0.195	0.072	0.079	0.143	0.190
Panel H: Detention						
Placed in a closed educational centre	0.175***	0.171***	-0.031*	-0.014	-0.022	-0.003
	(0.015)	(0.018)	(0.016)	(0.019)	(0.014)	(0.015)
Counterfactual mean	0.246	0.291	0.269	0.295	0.678	0.714
Observations	20,444	13,217	20,444	13,217	20,444	13,217
Treated observations	8,381	6,196	8,381	6,196	8,381	6,196
Adjusted R^2	0.180	0.224	0.121	0.146	0.105	0.159
Included in all regressions:						
Pre-spell judicial history and demographic controls	Y	Y	Y	Y	Y	Y
Year fixed effects	Y	Y	Y	Y	Y	Y
Department fixed effects	Y	Y	Y	Y	Y	Y

Notes: This table reports weighted least squares estimates of the average effect of a first closed educational centre placement on the treated. Each panel restricts the control group to the indicated type of alternative measure; unless stated otherwise the treated group is identical across panels and the entropy-balancing weights are re-estimated within each panel so that the restricted control pool matches the treated covariate profile (weighting method by panel: All eligible controls: Entropy (base weights); Community supervision (all): Entropy (base weights); Judicial supervision only: Entropy (base weights); Pre-trial supervised release: Entropy (base weights); Suspended sentence / supervised release: Entropy (base weights); Residential placement (excl. CER): Entropy (base weights); Reinforced educational centre (CER): Entropy (base weights); Detention: Entropy (base weights)). Each column corresponds to one post-spell outcome and horizon; outcomes at twelve months are defined only for spells whose full one-year post-spell window is observable before the data end and the minor's eighteenth birthday, and therefore over-represent minors placed young. The counterfactual mean is the weighted control mean of the outcome. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, in parentheses, are clustered at the prescriber-department level. Every comparison spell of an untreated minor receives a base weight equal to the inverse of the number of comparison spells that minor contributes to the panel, counted within each panel's own comparison pool, so that each untreated minor carries the same total base weight regardless of the length of the trajectory. The balancing weights are then estimated on top of these base weights, so that exact covariate balance with the treated group is re-imposed on the down-weighted pool. Treated spells are unique by minor by construction. Outcomes are anchored at the decision date over fixed windows of twelve and eighteen months. Everything else is identical to the main specification. The estimates on subsequent detention rise under this weighting, while the effects on a new closed centre placement and on a new court referral are more sensitive to it. Source: SSER data [GAME/IMAGES, 2012-2022].

Table C11: Effect of a first CEF placement with pre-decision outcomes in the conditioning set

	Detention episode		New CEF placement		New court referral	
	12 months	Up to age 18	12 months	Up to age 18	12 months	Up to age 18
Panel A: All eligible controls						
Placed in a closed educational centre	0.132***	0.179***	0.094***	0.115***	0.095***	0.155***
	(0.010)	(0.007)	(0.012)	(0.010)	(0.009)	(0.006)
Counterfactual mean	0.279	0.257	0.157	0.121	0.563	0.505
Observations	113,750	216,867	113,750	216,867	113,750	216,867
Treated observations	6,293	10,250	6,293	10,250	6,293	10,250
Adjusted R^2	0.166	0.124	0.073	0.086	0.125	0.111
Panel B: Community supervision (all)						
Placed in a closed educational centre	0.161***	0.265***	0.151***	0.177***	0.105***	0.257***
	(0.011)	(0.008)	(0.012)	(0.011)	(0.011)	(0.008)
Counterfactual mean	0.218	0.179	0.095	0.065	0.513	0.418
Observations	68,163	139,650	68,163	139,650	68,163	139,650
Treated observations	6,293	10,250	6,293	10,250	6,293	10,250
Adjusted R^2	0.158	0.117	0.078	0.081	0.127	0.096
Panel C: Judicial supervision only						
Placed in a closed educational centre	0.108***	0.196***	0.118***	0.149***	0.024*	0.134***
	(0.011)	(0.010)	(0.013)	(0.011)	(0.013)	(0.010)
Counterfactual mean	0.303	0.250	0.139	0.093	0.624	0.529
Observations	18,539	37,139	18,539	37,139	18,539	37,139
Treated observations	6,293	10,250	6,293	10,250	6,293	10,250
Adjusted R^2	0.130	0.134	0.103	0.122	0.129	0.112
Panel D: Pre-trial supervised release						
Placed in a closed educational centre	0.185***	0.288***	0.165***	0.175***	0.097***	0.266***
	(0.020)	(0.012)	(0.018)	(0.015)	(0.017)	(0.011)
Counterfactual mean	0.202	0.141	0.087	0.049	0.544	0.380
Observations	36,697	61,511	36,697	61,511	36,697	61,511
Treated observations	6,293	10,250	6,293	10,250	6,293	10,250
Adjusted R^2	0.157	0.205	0.108	0.158	0.125	0.210
Panel E: Suspended sentence / supervised release						
Placed in a closed educational centre	0.164***	0.305***	0.188***	0.211***	0.150***	0.343***
	(0.015)	(0.010)	(0.017)	(0.013)	(0.015)	(0.009)
Counterfactual mean	0.211	0.156	0.073	0.045	0.461	0.355
Observations	14,763	30,346	14,763	30,346	14,763	30,346
Treated observations	6,293	10,250	6,293	10,250	6,293	10,250
Adjusted R^2	0.192	0.192	0.121	0.146	0.165	0.184
Panel F: Residential placement (excl. CER)						
Placed in a closed educational centre	0.104***	0.096***	0.023*	0.025**	0.048***	0.033***
	(0.011)	(0.009)	(0.014)	(0.011)	(0.012)	(0.009)
Counterfactual mean	0.313	0.335	0.231	0.204	0.618	0.616
Observations	28,774	46,548	28,774	46,548	28,774	46,548
Treated observations	6,293	10,250	6,293	10,250	6,293	10,250
Adjusted R^2	0.123	0.139	0.073	0.132	0.096	0.158
Panel G: Reinforced educational centre (CER)						
Placed in a closed educational centre	0.104***	0.065***	-0.007	-0.008	0.051***	0.008
	(0.014)	(0.010)	(0.019)	(0.015)	(0.013)	(0.008)
Counterfactual mean	0.314	0.344	0.252	0.224	0.613	0.621
Observations	11,879	18,806	11,879	18,806	11,879	18,806
Treated observations	6,293	10,250	6,293	10,250	6,293	10,250
Adjusted R^2	0.126	0.156	0.066	0.136	0.112	0.175
Panel H: Detention						
Placed in a closed educational centre	0.008	-0.002	0.054***	0.054***	-0.015	-0.015
	(0.014)	(0.011)	(0.014)	(0.012)	(0.011)	(0.010)
Counterfactual mean	0.409	0.420	0.171	0.156	0.668	0.645
Observations	14,468	28,400	14,468	28,400	14,468	28,400
Treated observations	6,293	10,250	6,293	10,250	6,293	10,250
Adjusted R^2	0.144	0.148	0.084	0.120	0.103	0.144
Included in all regressions:						
Pre-spell judicial history, demographic and pre-decision outcome controls	Y	Y	Y	Y	Y	Y
Year fixed effects	Y	Y	Y	Y	Y	Y
Department fixed effects	Y	Y	Y	Y	Y	Y

Notes: This table reports weighted least squares estimates of the average effect of a first closed educational centre placement on the treated. Each panel restricts the control group to the indicated type of alternative measure; unless stated otherwise the treated group is identical across panels and the entropy-balancing weights are re-estimated within each panel so that the restricted control pool matches the treated covariate profile (weighting method by panel: All eligible controls: Entropy; Community supervision (all): Entropy; Judicial supervision only: Entropy; Pre-trial supervised release: IPW; Suspended sentence / supervised release: Entropy; Residential placement (excl. CER): Entropy; Reinforced educational centre (CER): Entropy; Detention: Entropy). Each column corresponds to one post-spell outcome and horizon; outcomes at twelve months are defined only for spells whose full one-year post-spell window is observable before the data end and the minor's eighteenth birthday, and therefore over-represent minors placed young, while outcomes up to age eighteen accumulate events over heterogeneous windows. The counterfactual mean is the weighted control mean of the outcome. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, in parentheses, are clustered at the prescriber-department level. Relative to the published specification, the balancing set adds the recency of the last judicial contact before the decision and the detention and court-referral rates over the twelve months preceding it. These are the three variables on which the falsification test reports residual imbalance. Nothing else changes: the common-support rule, the outcome definitions, the fixed effects and the clustering are identical. Source: SSER data [GAME/IMAGES, 2012-2022].

Table C12: Concentration of the entropy-balancing weights, by comparison panel

Panel	Control spells	Effective sample size	Max weight	p95	p99	Share of weight in top 1%
All eligible controls	339622	165709	27.02	2.82	5.22	7.3%
Community supervision (all)	238030	80590	53.43	3.02	6.81	10.7%
Judicial supervision only	53576	18119	49.24	2.84	6.46	10.8%
Pre-trial supervised release	73958	6435	3.76	0.44	3.76	27.1%
Suspended sentence / supervised release	40688	19509	32.11	2.65	5.22	7.8%
Residential placement (excl. CER)	43301	30224	12.23	2.23	3.56	4.5%
Reinforced educational centre (CER)	9731	7232	8.23	2.05	3.29	4.3%
Detention	23237	9690	27.65	2.90	5.88	8.5%

Notes: This table reports how concentrated the reweighting is within each comparison panel. The effective sample size is the squared sum of the control weights over the sum of their squares, and is the number of equally weighted control spells that would carry the same information. A panel whose effective sample size is far below its number of control spells attains covariate balance through a small number of heavily weighted comparisons. Source: SSER data [GAME/IMAGES, 2012-2022].

D The Placement Decision

Table D1: Judicial response as a choice between alternatives and as a position on a single severity dimension

	Multinomial: relative risk ratio against community supervision			Ordered
Covariate	Residential placement	Closed educational centre	Detention	Severity
Prior judicial episodes (any)	1.03***	1.05***	1.75***	1.32***
	(0.007)	(0.011)	(0.007)	(0.005)
Prior criminal court referrals	0.97***	0.97***	0.94***	0.96***
	(0.005)	(0.009)	(0.006)	(0.004)
Prior foyer placements	1.93***	1.84***	1.50***	1.68***
	(0.034)	(0.063)	(0.053)	(0.027)
Prior reinforced centre placements	1.58***	2.38***	1.58***	1.66***
	(0.038)	(0.069)	(0.058)	(0.030)
Prior judicial supervision orders	0.77***	1.02	0.55***	0.71***
	(0.013)	(0.021)	(0.014)	(0.009)
Prior suspended sentences with probation	0.49***	0.48***	0.48***	0.55***
	(0.029)	(0.052)	(0.029)	(0.021)
Prior pre-trial supervised releases	1.26***	1.11***	0.59***	0.89***
	(0.010)	(0.019)	(0.013)	(0.008)
Prior supervised releases	1.05***	1.05*	0.63***	0.83***
	(0.015)	(0.027)	(0.019)	(0.011)
Prior reparation measures	0.95***	0.96***	0.49***	0.71***
	(0.009)	(0.016)	(0.011)	(0.007)
Prior article 16bis measures	1.37***	1.13***	0.65***	0.96***
	(0.014)	(0.025)	(0.017)	(0.010)
Prior sentence adjustments	0.68***	0.77***	0.41***	0.56***
	(0.026)	(0.045)	(0.031)	(0.019)
Prior educational sanctions	0.78***	0.93	0.43***	0.62***
	(0.029)	(0.053)	(0.038)	(0.023)
Prior daytime educational activities	0.74***	0.60***	0.45***	0.57***
	(0.098)	(0.181)	(0.117)	(0.074)
Prior investigation measures	0.95***	0.99	0.58***	0.77***
	(0.008)	(0.012)	(0.008)	(0.005)
Age at the indexed spell	0.78***	0.60***	1.10***	0.85***
	(0.006)	(0.013)	(0.009)	(0.005)
Age at first judicial contact	1.09***	1.11***	1.15***	1.10***
	(0.006)	(0.013)	(0.008)	(0.005)
Active judicial supervision order	3.12***	4.18***	2.93***	2.88***
	(0.016)	(0.028)	(0.019)	(0.012)
Active suspended sentence	2.91***	3.47***	3.21***	2.75***
	(0.036)	(0.064)	(0.036)	(0.026)
Male	0.85***	1.22***	1.27***	0.99
	(0.016)	(0.040)	(0.028)	(0.013)
Prior detention episodes	1.15***	1.54***	1.12***	1.27***
	(0.014)	(0.023)	(0.015)	(0.010)
Prior placements (any type)	0.97	0.82***	0.45***	0.66***
	(0.033)	(0.061)	(0.051)	(0.026)
First observed spell before 2010	0.77***	0.94	1.02	0.86***
	(0.034)	(0.073)	(0.043)	(0.027)

Notes: This table compares two readings of the same decision, estimated on the 2012-2022 working sample with the same pre-spell covariates, a truncated-history cohort indicator and year dummies. The first three columns treat the four response families as distinct alternatives and report relative risk ratios against community supervision: a ratio above one means the covariate shifts the minor towards that alternative rather than community supervision, and each covariate carries its own ratio for each alternative. The last column treats the same four families as ordered positions on a single severity dimension, from community supervision to detention, so that each covariate carries one coefficient governing movement along that dimension; it is reported on the same scale, a value above one shifting the minor towards more severe responses. Comparing the two shows whether a covariate acts in one direction throughout, which the ordered reading requires, or in opposite directions across alternatives, which it cannot represent. McFadden pseudo R-squared: 0.116 for the multinomial and 0.084 for the ordered model. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Maximum-likelihood standard errors of the log coefficients in parentheses. Count covariates are capped at their 99th percentile. Source: SSER data [GAME/IMAGES, 2012-2022].

Table D2: Rank of the receiving facility in the distance ordering of centres

Rank	Placements	Share	Cumulative	Mean distance (km)	Standard deviation
1	2,862	23.4%	23.4%	35	33
2	1,934	15.8%	39.2%	65	40
3	1,493	12.2%	51.4%	87	48
4	1,161	9.5%	60.9%	100	56
5	797	6.5%	67.5%	131	57
6 to 10	1,817	14.9%	82.3%	183	61
Beyond 10	2,161	17.7%	100.0%	330	89

Notes:
This table ranks, for each first closed educational centre placement, all centres open in the year of the placement by their distance from the prescribing court, and reports the rank of the centre that actually admitted the minor. The sample covers metropolitan placements whose prescribing court is geolocated and whose receiving centre is matched to the official registry (12,225 placements). On average 49 centres are open in a given year; the median rank of the receiving centre is 3 and the median distance travelled is 100 kilometres, against 27 kilometres for placements in the nearest centre. The mean distance is 133 kilometres with a standard deviation of 128 kilometres, and 77.8 per cent of placements are executed in a department other than the prescribing one. Distances are measured from the prescribing court rather than from the minor's home, which the data do not record, so they measure how far the matching travelled within the judicial map rather than the distance actually separating the minor from their family. Source: SSER data [GAME/IMAGES, 2012-2022].

Table D3: Matching of minors to centres before and after the 2016 shift to a national admission procedure

Period	Placements	Share sent to the nearest centre	Median rank	Mean distance (km)	Share placed outside the department
Before 2016	3,388	24.9%	3	134	78.0%
2016 and after	8,837	22.8%	3	132	77.7%

Notes: This table compares how minors were matched to closed educational centres before and after the 2016 circular, which replaced a locally organised search for a place by a national procedure. Placements are ranked by the distance from the prescribing court to every centre open that year, so that rank one is the nearest centre. If proximity governed the match while the procedure was local and stopped doing so once it became national, the share sent to the nearest centre would fall and the distance travelled would rise across the two periods. Differences between periods are tested on the underlying placement-level distributions: the rank and the distance by a rank-sum test, the share sent to the nearest centre by a test of proportions, giving p-values of 0.215, 0.032 and 0.016 respectively. The sample is the same as that of the ranking table, namely metropolitan placements whose prescribing court is geolocated and whose receiving centre is matched to the official registry. Source: SSER data [GAME/IMAGES, 2012-2022].

Table D4: Occupancy of the skipped nearer centres at the month of the placement

	Placements	Occupancy of the receiving centre	Occupancy of the nearer centres not retained	Difference
Rank 2 to 3	3,291	1.055	1.027	-0.027
Rank 4 to 10	3,772	1.031	1.012	-0.019
Beyond rank 10	2,159	1.050	1.013	-0.037

Notes: This table compares, in the month of each placement, the occupancy of the centre that admitted the minor with the average occupancy of the centres that were closer to the prescribing court and did not receive the placement. Occupancy is the number of administratively open placement spells in the centre that month divided by its authorised capacity, capped at two; it measures recorded rather than physical presence, and the frozen-place phenomenon documented elsewhere in this project means it overstates true occupancy. A positive difference indicates that the nearer centres were more heavily loaded than the one finally retained, which is the signature of a capacity-constrained match; a difference near zero indicates that distance ranking was not resolved by availability. The first row is mechanically empty of skipped centres and is reported for completeness. The comparison is descriptive and conditions on the placement having taken place. Source: SSER data [GAME/IMAGES, 2012-2022].

Table D5: Local availability of a place and the probability that a recommended minor is actually placed

Term	(1)	(2)	(3)
Centres open in the court's department	0.1215***	0.1138***	
	(0.0214)	(0.0271)	
Authorised places in the court's department			0.0095***
			(0.0023)
Observations	4,302	4,302	4,302
Demographic controls	No	Yes	Yes
Year fixed effects	No	Yes	Yes

Notes: This table asks what separates recommended minors who are placed in a closed educational centre from recommended minors who are redirected elsewhere. The sample is the set of penal measures whose attached report recommended a closed centre; the dependent variable equals one when the placement actually took place. The explanatory variables measure how much capacity was open in the department of the competent court in the year of the decision, first as a count of centres and then as a count of authorised places. A positive coefficient means that recommended minors are more likely to be placed where a place exists, so that part of the departure from the recommendation reflects availability rather than a judgement on the case. A coefficient close to zero leaves the departure unexplained by capacity, and the comparison group is then more likely to differ from the placed group on the severity of the case, in which case the estimated effect should be read as an upper bound. Standard errors are clustered at the court-department level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Source: SSER data [PARCOURS, 2021-2024].

Table D6: Facilities ranked by the severity of the profiles they admit

	Facilities	Placements	Severity index	Share with a prior detention	Mean prior episodes	Mean age	Share publicly run
Lightest quarter of facilities	13	3,000	-0.15	0.081	7.0	15.1	0.38
Second quarter	13	3,125	-0.03	0.183	7.3	16.1	0.15
Third quarter	12	3,403	0.04	0.282	7.7	16.7	0.42
Heaviest quarter of facilities	13	3,291	0.12	0.305	8.0	16.7	0.38

Notes: This table groups closed educational centres into quarters according to the severity of the minors they admit, and describes each quarter. The severity index averages three standardised pre-placement histories - prior judicial episodes, prior placements and prior detention episodes - after removing prescribing-department-by-year averages, so that facilities are compared on the minors sent by the same courts in the same year; it is centred on zero by construction. The sample covers first placements in the 51 facilities with at least thirty placements over 2012-2022 (12,819 placements). A facility in the heaviest quarter admits minors whose recorded history is markedly heavier than the minors sent by the same courts in the same years to other facilities. Source: SSER data [GAME/IMAGES, 2012-2022].

Table D7: Year-to-year transitions of facilities across thirds of intake severity

	Position in year $t+1$		
Position in year t	Lightest	Middle	Heaviest
Lightest	47.4%	31.6%	21.1%
Middle	31.1%	33.8%	35.1%
Heaviest	22.4%	34.0%	43.6%

Notes: Each cell reports the share of facility-years in the row position that are in the column position the following year, over 456 consecutive facility-year pairs. Facilities are ranked each year by the average severity of the profiles they admit, net of prescribing-department-by-year averages, and split into thirds; facility-years with fewer than ten placements are dropped. Under random matching every cell would be close to 33 percent: the diagonal measures how much a facility's position persists from one year to the next. Source: SSER data [GAME/IMAGES, 2012-2022].

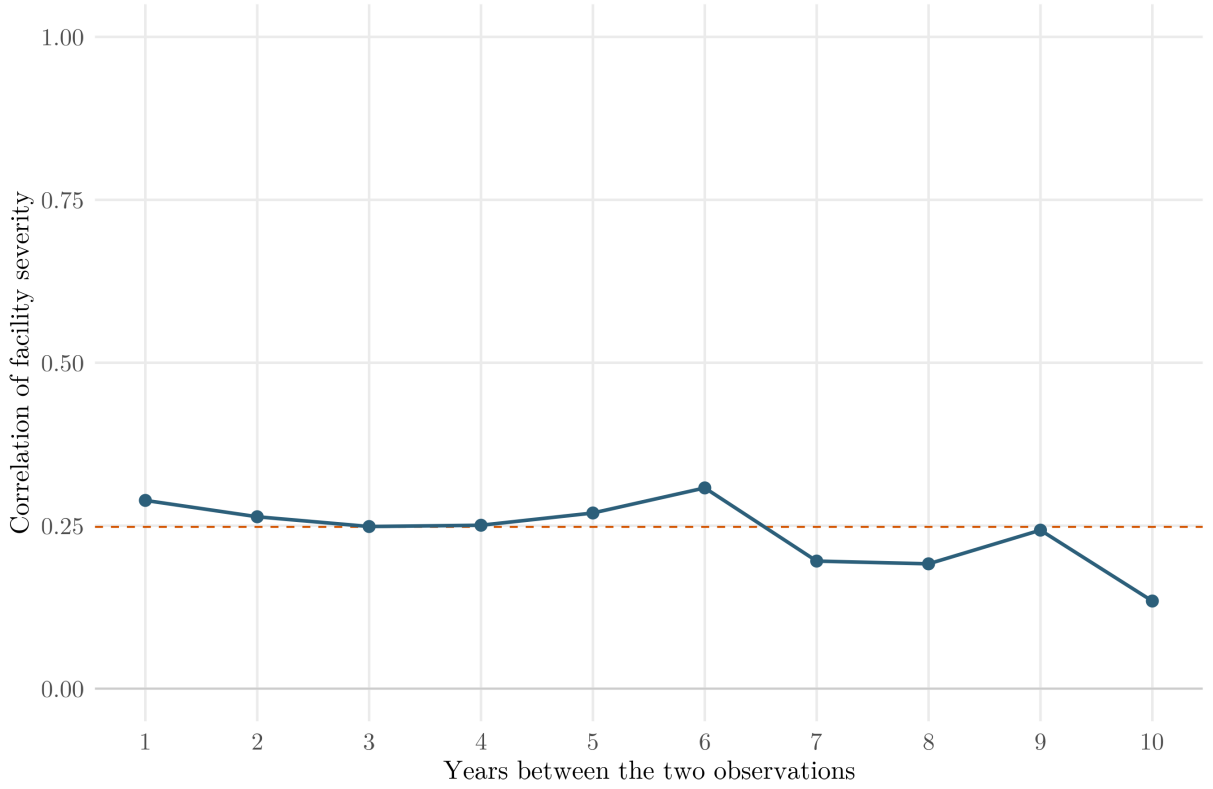

Figure D1: Persistence of facility intake profiles across years

Notes: The figure plots the average correlation, across all pairs of calendar years separated by the indicated number of years, between facilities' average severity of admitted profiles in the first year and in the second. Severity averages three standardised pre-placement histories net of prescribing-department-by-year averages, so facilities are compared on minors sent by the same courts in the same years; facility-years with fewer than ten placements are dropped. The dashed line at 0.25 is the share of the dispersion between facilities that is permanent, estimated as the average covariance between distinct years over the average variance within a year - a measure that sampling noise does not contaminate, since that noise is independent across years. Source: SSER data [GAME/IMAGES, 2012-2022].

Table D8: Rank and distance of the receiving facility by severity of the minor's history

	Rank of the receiving facility	Distance travelled (km)
Severity of the minor's history	0.529***	8.38***
	(0.163)	(2.47)
Observations	11,747	11,747
Court-by-year fixed effects	Y	Y

Notes: This table regresses how far down the distance ranking the receiving facility sits, and how many kilometres the minor travels, on the severity of the minor's own pre-placement history, with prescribing-department-by-year fixed effects and standard errors clustered at the facility level. Severity averages three standardised pre-placement histories. A positive coefficient means that, among minors sent by the same courts in the same year, those with heavier records are placed in facilities that are further down the list of nearby options - the pattern that refusals by nearer facilities would generate. The estimate is descriptive and conditions on a placement having taken place. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Source: SSER data [GAME/IMAGES, 2012-2022].

E Additional Estimates

Table E1: Effect of a first CEF placement by counterfactual type, comparison-specific support

	Detention episode		New CEF placement		New court referral	
	12 months	18 months	12 months	18 months	12 months	18 months
Panel A: Community supervision (all)						
Placed in a closed educational centre	0.214***	0.227***	0.153***	0.173***	0.197***	0.171***
	(0.011)	(0.011)	(0.011)	(0.013)	(0.009)	(0.009)
Counterfactual mean	0.199	0.229	0.122	0.155	0.448	0.535
Observations	183,447	141,416	183,447	141,416	183,447	141,416
Treated observations	8,374	6,186	8,374	6,186	8,374	6,186
Adjusted R^2	0.212	0.233	0.100	0.124	0.177	0.181
Panel B: Judicial supervision only						
Placed in a closed educational centre	0.150***	0.157***	0.092***	0.107***	0.099***	0.072***
	(0.011)	(0.012)	(0.011)	(0.013)	(0.010)	(0.011)
Counterfactual mean	0.267	0.316	0.182	0.224	0.537	0.636
Observations	42,676	32,467	42,676	32,467	42,676	32,467
Treated observations	8,298	6,148	8,298	6,148	8,298	6,148
Adjusted R^2	0.191	0.214	0.106	0.129	0.145	0.160
Panel C: Pre-trial supervised release						
Placed in a closed educational centre	0.255***	0.272***	0.141***	0.143***	0.165***	0.144***
	(0.019)	(0.017)	(0.017)	(0.018)	(0.013)	(0.015)
Counterfactual mean	0.213	0.250	0.130	0.182	0.524	0.613
Observations	35,558	27,469	35,558	27,469	35,558	27,469
Treated observations	7,801	5,583	7,801	5,583	7,801	5,583
Adjusted R^2	0.200	0.233	0.115	0.139	0.166	0.198
Panel D: Suspended sentence / supervised release						
Placed in a closed educational centre	0.205***	0.212***	0.169***	0.195***	0.225***	0.205***
	(0.012)	(0.013)	(0.013)	(0.016)	(0.011)	(0.011)
Counterfactual mean	0.210	0.249	0.111	0.140	0.436	0.520
Observations	34,309	25,449	34,309	25,449	34,309	25,449
Treated observations	8,188	6,093	8,188	6,093	8,188	6,093
Adjusted R^2	0.227	0.268	0.116	0.143	0.185	0.203
Panel E: Residential placement (excl. CER)						
Placed in a closed educational centre	0.148***	0.151***	0.056***	0.061***	0.067***	0.044***
	(0.011)	(0.011)	(0.013)	(0.015)	(0.011)	(0.011)
Counterfactual mean	0.284	0.336	0.233	0.292	0.604	0.701
Observations	35,005	25,663	35,005	25,663	35,005	25,663
Treated observations	8,196	6,023	8,196	6,023	8,196	6,023
Adjusted R^2	0.155	0.184	0.075	0.085	0.098	0.123
Panel F: Reinforced educational centre (CER)						
Placed in a closed educational centre	0.166***	0.151***	0.066***	0.050**	0.121***	0.102***
	(0.014)	(0.016)	(0.016)	(0.022)	(0.011)	(0.011)
Counterfactual mean	0.258	0.325	0.213	0.298	0.542	0.630
Observations	13,952	9,955	13,952	9,955	13,952	9,955
Treated observations	8,078	5,905	8,078	5,905	8,078	5,905
Adjusted R^2	0.169	0.194	0.071	0.077	0.147	0.192
Panel G: Detention						
Placed in a closed educational centre	0.113***	0.099***	0.025*	0.039**	-0.049***	-0.032**
	(0.014)	(0.017)	(0.015)	(0.018)	(0.014)	(0.013)
Counterfactual mean	0.308	0.372	0.221	0.251	0.708	0.753
Observations	20,534	13,301	20,534	13,301	20,534	13,301
Treated observations	8,503	6,219	8,503	6,219	8,503	6,219
Adjusted R^2	0.168	0.202	0.099	0.123	0.102	0.141
Included in all regressions:						
Pre-spell judicial history and demographic controls	Y	Y	Y	Y	Y	Y
Year fixed effects	Y	Y	Y	Y	Y	Y
Department fixed effects	Y	Y	Y	Y	Y	Y

Notes: This table reports weighted least squares estimates of the average effect of a first closed educational centre placement on the treated. Each panel restricts the control group to the indicated type of alternative measure; unless stated otherwise the treated group is identical across panels and the entropy-balancing weights are re-estimated within each panel so that the restricted control pool matches the treated covariate profile (weighting method by panel: Community supervision (all): Entropy; Judicial supervision only: Entropy; Pre-trial supervised release: IPW; Suspended sentence / supervised release: Entropy; Residential placement (excl. CER): Entropy; Reinforced educational centre (CER): Entropy; Detention: Entropy). Each column corresponds to one post-spell outcome and horizon; outcomes at twelve months are defined only for spells whose full one-year post-spell window is observable before the data end and the minor's eighteenth birthday, and therefore over-represent minors placed young. The counterfactual mean is the weighted control mean of the outcome. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, in parentheses, are clustered at the prescriber-department level. In this table the propensity score, the common support (tenth percentile of the within-comparison treated scores) and the weights are re-estimated within each comparison. The treated population therefore differs across panels - each panel retains the treated spells comparable to that specific alternative - and coefficients are not comparable across panels. Source: SSER data [GAME/IMAGES, 2012-2022].

Table E2: Effect of a first CEF placement within the legal vehicle of the placement

	Detention episode		New CEF placement		New court referral	
	12 months	18 months	12 months	18 months	12 months	18 months
Panel A: CEF under judicial supervision vs judicial supervision alone						
Placed in a closed educational centre	0.159***	0.177***	0.075***	0.086***	0.118***	0.107***
	(0.012)	(0.013)	(0.012)	(0.014)	(0.010)	(0.010)
Counterfactual mean	0.307	0.364	0.207	0.254	0.573	0.680
Observations	36,143	28,297	36,143	28,297	36,143	28,297
Treated observations	6,240	4,594	6,240	4,594	6,240	4,594
Adjusted R^2	0.177	0.199	0.112	0.133	0.137	0.151
Panel B: Supervision order already in force at the placement						
Placed in a closed educational centre	0.153***	0.171***	0.049***	0.060***	0.075***	0.059***
	(0.015)	(0.016)	(0.011)	(0.016)	(0.012)	(0.011)
Counterfactual mean	0.362	0.437	0.237	0.291	0.618	0.745
Observations	10,744	7,399	10,744	7,399	10,744	7,399
Treated observations	4,252	3,104	4,252	3,104	4,252	3,104
Adjusted R^2	0.137	0.146	0.098	0.114	0.109	0.103
Panel C: Supervision order made jointly with the placement						
Placed in a closed educational centre	0.195***	0.197***	0.128***	0.135***	0.223***	0.203***
	(0.014)	(0.014)	(0.017)	(0.019)	(0.015)	(0.015)
Counterfactual mean	0.188	0.235	0.141	0.181	0.474	0.570
Observations	34,728	25,556	34,728	25,556	34,728	25,556
Treated observations	2,146	1,574	2,146	1,574	2,146	1,574
Adjusted R^2	0.149	0.183	0.113	0.147	0.142	0.154
Panel D: CEF under suspended sentence vs suspended sentence alone						
Placed in a closed educational centre	0.209***	0.185***	0.090***	0.137***	0.171***	0.187***
	(0.028)	(0.046)	(0.024)	(0.037)	(0.023)	(0.029)
Counterfactual mean	0.395	0.497	0.170	0.208	0.611	0.698
Observations	3,680	2,419	3,680	2,419	3,680	2,419
Treated observations	1,103	716	1,103	716	1,103	716
Adjusted R^2	0.157	0.222	0.152	0.216	0.111	0.182
Included in all regressions:						
Pre-spell judicial history and demographic controls	Y	Y	Y	Y	Y	Y
Year fixed effects	Y	Y	Y	Y	Y	Y
Department fixed effects	Y	Y	Y	Y	Y	Y

Notes: This table reports weighted least squares estimates of the average effect of a first closed educational centre placement on the treated. Each panel restricts the control group to the indicated type of alternative measure; unless stated otherwise the treated group is identical across panels and the entropy-balancing weights are re-estimated within each panel so that the restricted control pool matches the treated covariate profile (weighting method by panel: CEF under judicial supervision vs judicial supervision alone: Entropy; Supervision order already in force at the placement: Entropy; Supervision order made jointly with the placement: Entropy; CEF under suspended sentence vs suspended sentence alone: Entropy). Each column corresponds to one post-spell outcome and horizon; outcomes at twelve months are defined only for spells whose full one-year post-spell window is observable before the data end and the minor's eighteenth birthday, and therefore over-represent minors placed young. The counterfactual mean is the weighted control mean of the outcome. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, in parentheses, are clustered at the prescriber-department level. Each panel restricts the treated group to first CEF placements whose accompanying coercive vehicle (an order active at the indexed spell or prescribed within seven days of it) matches the control group's measure, and the control group to standalone spells of that measure. The propensity score, the support and the weights are estimated within each comparison. The second and third panels split the treated group of the first according to whether the supervision order was already running when the placement was ordered or was made at the same time, against the same comparison group of minors under a standalone order. An order created on the occasion of the placement indicates that the coercive vehicle was chosen in order to carry it, whereas an order already in force means the placement was added to a constraint the minor was already under. Source: SSER data [GAME/IMAGES, 2012-2022].

Table E3: Effect of a first CEF placement by counterfactual type, retaining single-record placements

	Detention episode		New CEF placement		New court referral	
	12 months	18 months	12 months	18 months	12 months	18 months
Panel A: All eligible controls						
Placed in a closed educational centre	0.186***	0.193***	0.119***	0.134***	0.141***	0.114***
	(0.010)	(0.010)	(0.010)	(0.012)	(0.008)	(0.009)
Counterfactual mean	0.203	0.236	0.138	0.173	0.465	0.545
Observations	273,674	200,528	273,674	200,528	273,674	200,528
Treated observations	9,113	6,720	9,113	6,720	9,113	6,720
Adjusted R^2	0.197	0.227	0.100	0.124	0.185	0.197
Panel B: Community supervision (all)						
Placed in a closed educational centre	0.197***	0.205***	0.141***	0.157***	0.170***	0.134***
	(0.011)	(0.010)	(0.011)	(0.013)	(0.009)	(0.009)
Counterfactual mean	0.187	0.218	0.114	0.147	0.432	0.522
Observations	200,606	148,885	200,606	148,885	200,606	148,885
Treated observations	9,113	6,720	9,113	6,720	9,113	6,720
Adjusted R^2	0.213	0.232	0.101	0.123	0.176	0.178
Panel C: Judicial supervision only						
Placed in a closed educational centre	0.137***	0.137***	0.081***	0.091***	0.077***	0.038***
	(0.010)	(0.011)	(0.011)	(0.013)	(0.010)	(0.011)
Counterfactual mean	0.253	0.305	0.171	0.216	0.519	0.625
Observations	50,545	35,711	50,545	35,711	50,545	35,711
Treated observations	9,113	6,720	9,113	6,720	9,113	6,720
Adjusted R^2	0.195	0.216	0.111	0.134	0.151	0.166
Panel D: Pre-trial supervised release						
Placed in a closed educational centre	0.226***	0.232***	0.155***	0.163***	0.154***	0.113***
	(0.016)	(0.014)	(0.014)	(0.015)	(0.012)	(0.014)
Counterfactual mean	0.157	0.189	0.097	0.135	0.442	0.535
Observations	85,682	71,259	85,682	71,259	85,682	71,259
Treated observations	9,113	6,720	9,113	6,720	9,113	6,720
Adjusted R^2	0.226	0.262	0.122	0.145	0.201	0.218
Panel E: Suspended sentence / supervised release						
Placed in a closed educational centre	0.191***	0.194***	0.154***	0.178***	0.199***	0.171***
	(0.011)	(0.012)	(0.012)	(0.014)	(0.010)	(0.011)
Counterfactual mean	0.198	0.239	0.105	0.134	0.422	0.507
Observations	40,920	28,933	40,920	28,933	40,920	28,933
Treated observations	9,113	6,720	9,113	6,720	9,113	6,720
Adjusted R^2	0.231	0.271	0.116	0.142	0.185	0.202
Panel F: Residential placement (excl. CER)						
Placed in a closed educational centre	0.131***	0.131***	0.046***	0.048***	0.037***	0.006
	(0.011)	(0.011)	(0.013)	(0.014)	(0.011)	(0.012)
Counterfactual mean	0.268	0.320	0.223	0.281	0.587	0.685
Observations	40,133	29,069	40,133	29,069	40,133	29,069
Treated observations	9,113	6,720	9,113	6,720	9,113	6,720
Adjusted R^2	0.162	0.191	0.079	0.091	0.110	0.140
Panel G: Reinforced educational centre (CER)						
Placed in a closed educational centre	0.152***	0.135***	0.057***	0.040**	0.093***	0.070***
	(0.012)	(0.014)	(0.015)	(0.020)	(0.010)	(0.010)
Counterfactual mean	0.245	0.310	0.207	0.289	0.531	0.617
Observations	16,420	11,793	16,420	11,793	16,420	11,793
Treated observations	9,113	6,720	9,113	6,720	9,113	6,720
Adjusted R^2	0.183	0.214	0.073	0.083	0.174	0.231
Panel H: Detention						
Placed in a closed educational centre	0.091***	0.074***	0.007	0.020	-0.089***	-0.079***
	(0.014)	(0.016)	(0.015)	(0.017)	(0.013)	(0.014)
Counterfactual mean	0.305	0.363	0.218	0.246	0.706	0.747
Observations	21,851	13,944	21,851	13,944	21,851	13,944
Treated observations	9,113	6,720	9,113	6,720	9,113	6,720
Adjusted R^2	0.173	0.211	0.103	0.128	0.118	0.162
Included in all regressions:						
Pre-spell judicial history and demographic controls	Y	Y	Y	Y	Y	Y
Year fixed effects	Y	Y	Y	Y	Y	Y
Department fixed effects	Y	Y	Y	Y	Y	Y

Notes: This table reports weighted least squares estimates of the average effect of a first closed educational centre placement on the treated. Each panel restricts the control group to the indicated type of alternative measure; unless stated otherwise the treated group is identical across panels and the entropy-balancing weights are re-estimated within each panel so that the restricted control pool matches the treated covariate profile (weighting method by panel: All eligible controls: Entropy; Community supervision (all): Entropy; Judicial supervision only: Entropy; Pre-trial supervised release: IPW; Suspended sentence / supervised release: Entropy; Residential placement (excl. CER): Entropy; Reinforced educational centre (CER): Entropy; Detention: Entropy). Each column corresponds to one post-spell outcome and horizon; outcomes at twelve months are defined only for spells whose full one-year post-spell window is observable before the data end and the minor's eighteenth birthday, and therefore over-represent minors placed young. The counterfactual mean is the weighted control mean of the outcome. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, in parentheses, are clustered at the prescriber-department level. This robustness specification reintroduces into the treated pool the first placements that constitute the minor's only observed record in the database; the common support and the weights are re-estimated accordingly. Source: SSER data [GAME/IMAGES, 2012-2022].

F Estimates Anchored at the End of the Measure

Table F1: Effect of a CEF placement within the recommended population (PARCOURS, 2021-2024)

	Detention episode		New CEF placement		New court referral	
	12 months	Up to age 18	12 months	Up to age 18	12 months	Up to age 18
Panel A: Educator recommendation sample						
Placed in a closed educational centre	0.037	0.093***	0.017	0.049***	-0.002	0.141***
	(0.033)	(0.018)	(0.029)	(0.017)	(0.016)	(0.019)
Counterfactual mean	0.530	0.284	0.263	0.114	0.940	0.589
Observations	893	4,285	893	4,285	893	4,285
Treated observations	201	569	201	569	201	569
Adjusted R^2	0.246	0.181	0.180	0.143	0.260	0.183
Panel B: Formal decision sample						
Placed in a closed educational centre	0.047	0.116***	0.048	0.101***	-0.023	0.149***
	(0.042)	(0.028)	(0.048)	(0.019)	(0.035)	(0.027)
Counterfactual mean	0.521	0.254	0.226	0.087	0.911	0.567
Observations	413	1,739	413	1,739	413	1,739
Treated observations	135	393	135	393	135	393
Adjusted R^2	0.260	0.234	0.201	0.161	0.416	0.228
Panel C: Judicial supervision as the comparison						
Placed in a closed educational centre	0.029	0.087***	0.036	0.046**	-0.007	0.104***
	(0.046)	(0.029)	(0.037)	(0.018)	(0.018)	(0.027)
Counterfactual mean	0.565	0.295	0.273	0.123	0.950	0.637
Observations	374	1,358	374	1,358	374	1,358
Treated observations	201	569	201	569	201	569
Adjusted R^2	0.201	0.207	0.160	0.176	0.221	0.250
Included in all regressions:						
Pre-spell judicial history and demographic controls	Y	Y	Y	Y	Y	Y
Year fixed effects	Y	Y	Y	Y	Y	Y
Department fixed effects	Y	Y	Y	Y	Y	Y

Notes: This table reports weighted least squares estimates of the average effect of a first closed educational centre placement on the treated. Each panel restricts the control group to the indicated type of alternative measure; unless stated otherwise the treated group is identical across panels and the entropy-balancing weights are re-estimated within each panel so that the restricted control pool matches the treated covariate profile (weighting method by panel: Educator recommendation sample: Entropy; Formal decision sample: Entropy; Judicial supervision as the comparison: Entropy). Each column corresponds to one post-spell outcome and horizon; outcomes at twelve months are defined only for spells whose full one-year post-spell window is observable before the data end and the minor's eighteenth birthday, and therefore over-represent minors placed young, while outcomes up to age eighteen accumulate events over heterogeneous windows. The counterfactual mean is the weighted control mean of the outcome. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, in parentheses, are clustered at the prescriber-department level. The sample is restricted to penal measures whose attached social-educational report recommended a closed educational centre placement (Panel A: educator narrative recommendation; Panel B: formal post-investigation decision). Treated units are measures effectively executed in a closed centre; controls are recommended measures redirected to any other response. Panel C narrows the control group of Panel A to recommended minors who received a judicial supervision order in place of the placement, so that both arms are held under the same coercive instrument and differ only in the closed residential component; it is the counterpart, on the recommended population, of the same comparison run on the larger registry. Controls include the seventeen demographic, family and prior-history covariates of the PARCOURS analysis; the department is that of the competent juvenile court. Detention outcomes rely on the PARCOURS records, whose annual volume of custodial episodes runs at roughly a quarter of the GAME registry over the overlapping years; both systems register custody through the same educational taking-charge, so the gap reflects the progressive roll-out of the newer system and the identification of custodial episodes through the receiving structure rather than a difference in scope. Custodial episodes are identified through the educational service that takes charge of every detained minor, so their recording does not depend on an ongoing measure and does not differ between placed and comparison minors. The GAME-based estimates, whose detention registry is complete over the whole window, return the same order of magnitude. Outcome windows run to the earlier of the data end and the eighteenth birthday; because the panel opens in 2019, the twelve-month sub-sample is small and skews towards younger minors and earlier cohorts. Source: SSER data [PARCOURS, 2021-2024].

Table F2: Effect of a first CEF placement by counterfactual type, common-support sample

	Detention episode		New CEF placement		New court referral	
	12 months	Up to age 18	12 months	Up to age 18	12 months	Up to age 18
Panel A: All eligible controls						
Placed in a closed educational centre	0.136***	0.188***	0.109***	0.127***	0.100***	0.166***
	(0.009)	(0.007)	(0.011)	(0.010)	(0.009)	(0.006)
Counterfactual mean	0.249	0.235	0.141	0.111	0.520	0.478
Observations	136,242	248,086	136,242	248,086	136,242	248,086
Treated observations	6,659	10,659	6,659	10,659	6,659	10,659
Adjusted R^2	0.179	0.124	0.081	0.085	0.157	0.115
Panel B: Community supervision (all)						
Placed in a closed educational centre	0.163***	0.266***	0.162***	0.183***	0.105***	0.256***
	(0.010)	(0.008)	(0.012)	(0.011)	(0.011)	(0.008)
Counterfactual mean	0.190	0.164	0.085	0.062	0.476	0.403
Observations	80,058	158,300	80,058	158,300	80,058	158,300
Treated observations	6,659	10,659	6,659	10,659	6,659	10,659
Adjusted R^2	0.160	0.112	0.080	0.081	0.136	0.094
Panel C: Judicial supervision only						
Placed in a closed educational centre	0.101***	0.194***	0.127***	0.156***	0.014	0.131***
	(0.011)	(0.009)	(0.013)	(0.010)	(0.012)	(0.010)
Counterfactual mean	0.291	0.242	0.134	0.090	0.609	0.519
Observations	19,791	39,261	19,791	39,261	19,791	39,261
Treated observations	6,659	10,659	6,659	10,659	6,659	10,659
Adjusted R^2	0.132	0.132	0.098	0.120	0.135	0.113
Panel D: Pre-trial supervised release						
Placed in a closed educational centre	0.176***	0.282***	0.175***	0.180***	0.086***	0.254***
	(0.018)	(0.012)	(0.017)	(0.014)	(0.017)	(0.011)
Counterfactual mean	0.185	0.135	0.080	0.048	0.517	0.377
Observations	43,282	71,547	43,282	71,547	43,282	71,547
Treated observations	6,659	10,659	6,659	10,659	6,659	10,659
Adjusted R^2	0.167	0.202	0.109	0.158	0.140	0.199
Panel E: Suspended sentence / supervised release						
Placed in a closed educational centre	0.163***	0.303***	0.194***	0.215***	0.148***	0.336***
	(0.012)	(0.010)	(0.016)	(0.013)	(0.015)	(0.009)
Counterfactual mean	0.180	0.141	0.065	0.043	0.419	0.340
Observations	16,188	32,640	16,188	32,640	16,188	32,640
Treated observations	6,659	10,659	6,659	10,659	6,659	10,659
Adjusted R^2	0.199	0.187	0.126	0.147	0.182	0.173
Panel F: Residential placement (excl. CER)						
Placed in a closed educational centre	0.102***	0.097***	0.032**	0.031***	0.049***	0.037***
	(0.010)	(0.009)	(0.013)	(0.010)	(0.012)	(0.009)
Counterfactual mean	0.298	0.324	0.224	0.201	0.591	0.599
Observations	30,748	48,939	30,748	48,939	30,748	48,939
Treated observations	6,659	10,659	6,659	10,659	6,659	10,659
Adjusted R^2	0.132	0.139	0.074	0.129	0.115	0.156
Panel G: Reinforced educational centre (CER)						
Placed in a closed educational centre	0.109***	0.074***	0.005	0.001	0.074***	0.031***
	(0.012)	(0.010)	(0.018)	(0.014)	(0.012)	(0.008)
Counterfactual mean	0.291	0.328	0.242	0.220	0.564	0.588
Observations	12,931	19,978	12,931	19,978	12,931	19,978
Treated observations	6,659	10,659	6,659	10,659	6,659	10,659
Adjusted R^2	0.145	0.159	0.066	0.133	0.150	0.175
Panel H: Detention						
Placed in a closed educational centre	0.015	0.015	0.069***	0.069***	-0.015	-0.001
	(0.012)	(0.010)	(0.013)	(0.011)	(0.012)	(0.011)
Counterfactual mean	0.375	0.390	0.152	0.140	0.632	0.614
Observations	16,201	30,430	16,201	30,430	16,201	30,430
Treated observations	6,659	10,659	6,659	10,659	6,659	10,659
Adjusted R^2	0.161	0.145	0.091	0.117	0.134	0.141
Included in all regressions:						
Pre-spell judicial history and demographic controls	Y	Y	Y	Y	Y	Y
Year fixed effects	Y	Y	Y	Y	Y	Y
Department fixed effects	Y	Y	Y	Y	Y	Y

Notes: This table reports weighted least squares estimates of the average effect of a first closed educational centre placement on the treated. Each panel restricts the control group to the indicated type of alternative measure; unless stated otherwise the treated group is identical across panels and the entropy-balancing weights are re-estimated within each panel so that the restricted control pool matches the treated covariate profile (weighting method by panel: All eligible controls: Entropy; Community supervision (all): Entropy; Judicial supervision only: Entropy; Pre-trial supervised release: IPW; Suspended sentence / supervised release: Entropy; Residential placement (excl. CER): Entropy; Reinforced educational centre (CER): Entropy; Detention: Entropy). Each column corresponds to one post-spell outcome and horizon; outcomes at twelve months are defined only for spells whose full one-year post-spell window is observable before the data end and the minor's eighteenth birthday, and therefore over-represent minors placed young, while outcomes up to age eighteen accumulate events over heterogeneous windows. The counterfactual mean is the weighted control mean of the outcome. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, in parentheses, are clustered at the prescriber-department level. Against the pre-trial supervised release pool, exact entropy balance is not attained: that control pool carries lighter pre-spell histories than the treated group over the range where the two overlap. Source: SSER data [GAME/IMAGES, 2012-2022].

Table F3: Effect of a CEF placement by length of the stay, within the recommended population (PARCOURS, 2021-2024)

	Detention episode		New CEF placement		New court referral	
	12 months	Up to age 18	12 months	Up to age 18	12 months	Up to age 18
Panel A: Stay ending before four months						
Placed in a closed educational centre	0.086*	0.191***	0.082**	0.111***	-0.008	0.212***
	(0.044)	(0.025)	(0.040)	(0.023)	(0.027)	(0.018)
Counterfactual mean	0.551	0.292	0.270	0.119	0.936	0.585
Observations	825	4,048	825	4,048	825	4,048
Treated observations	133	332	133	332	133	332
Adjusted R^2	0.256	0.185	0.182	0.145	0.271	0.174
Panel B: Stay of four months or more						
Placed in a closed educational centre	-0.052	-0.039	-0.099**	-0.036*	0.013	0.036
	(0.071)	(0.040)	(0.045)	(0.020)	(0.024)	(0.031)
Counterfactual mean	0.499	0.271	0.253	0.108	0.948	0.595
Observations	760	3,953	760	3,953	760	3,953
Treated observations	68	237	68	237	68	237
Adjusted R^2	0.250	0.174	0.208	0.144	0.215	0.188
Included in all regressions:						
Pre-spell judicial history and demographic controls	Y	Y	Y	Y	Y	Y
Year fixed effects	Y	Y	Y	Y	Y	Y
Department fixed effects	Y	Y	Y	Y	Y	Y

Notes: This table reports weighted least squares estimates of the average effect of a first closed educational centre placement on the treated. Each panel restricts the control group to the indicated type of alternative measure; unless stated otherwise the treated group is identical across panels and the entropy-balancing weights are re-estimated within each panel so that the restricted control pool matches the treated covariate profile (weighting method by panel: Stay ending before four months: Entropy; Stay of four months or more: Entropy). Each column corresponds to one post-spell outcome and horizon; outcomes at twelve months are defined only for spells whose full one-year post-spell window is observable before the data end and the minor's eighteenth birthday, and therefore over-represent minors placed young, while outcomes up to age eighteen accumulate events over heterogeneous windows. The counterfactual mean is the weighted control mean of the outcome. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, in parentheses, are clustered at the prescriber-department level. The sample is restricted to penal measures whose attached social-educational report recommended a closed educational centre placement. Each panel keeps the same comparison group, namely recommended minors redirected to another response, and one of the two treated groups: placements whose recorded stay ended before four months in the first panel, placements of four months or more in the second. Entropy-balancing weights are re-estimated within each panel, so the comparison group is realigned on the covariate profile of the treated group it faces. Outcomes are measured from the end of the indexed measure, so the two treated groups are observed over windows of equal length. The length of a stay is itself an outcome of the placement and premature departure is associated with incidents during the stay and with revocation of the order, so the two treated groups are not random sub-samples of the placed population and the contrast between panels mixes a programme-completion effect with selection on how the stay unfolded. The comparison is descriptive. Source: SSER data [PARCOURS, 2021-2024].

Table F4: Placement effect on the subsample with a recorded offence nature (PARCOURS, 2021-2024)

	Detention episode		New CEF placement		New court referral	
	12 months	Up to age 18	12 months	Up to age 18	12 months	Up to age 18
Panel A: Covariate adjustment only						
Placed in a closed educational centre	0.104***	0.165***	0.105**	0.113***	0.029	0.197***
	(0.036)	(0.023)	(0.046)	(0.013)	(0.021)	(0.026)
Counterfactual mean	0.415	0.194	0.215	0.080	0.861	0.489
Observations	6,465	48,437	6,465	48,437	6,465	48,437
Treated observations	134	454	134	454	134	454
Adjusted R^2	0.305	0.156	0.203	0.116	0.181	0.155
Panel B: Adding the recommendation indicators						
Placed in a closed educational centre	0.082**	0.138***	0.102**	0.103***	0.025	0.182***
	(0.035)	(0.024)	(0.047)	(0.014)	(0.022)	(0.026)
Counterfactual mean	0.415	0.194	0.215	0.080	0.861	0.489
Observations	6,465	48,437	6,465	48,437	6,465	48,437
Treated observations	134	454	134	454	134	454
Adjusted R^2	0.307	0.160	0.207	0.120	0.181	0.156
Panel C: Adding the offence families as controls						
Placed in a closed educational centre	0.089**	0.142***	0.108**	0.105***	0.031	0.187***
	(0.035)	(0.024)	(0.043)	(0.014)	(0.024)	(0.026)
Counterfactual mean	0.415	0.194	0.215	0.080	0.861	0.489
Observations	6,465	48,437	6,465	48,437	6,465	48,437
Treated observations	134	454	134	454	134	454
Adjusted R^2	0.313	0.161	0.212	0.121	0.187	0.159
Panel D: Reweighting on the offence families as well						
Placed in a closed educational centre	0.095**	0.145***	0.105***	0.104***	0.033	0.186***
	(0.037)	(0.023)	(0.039)	(0.014)	(0.025)	(0.025)
Counterfactual mean	0.412	0.193	0.210	0.079	0.858	0.484
Observations	6,465	48,437	6,465	48,437	6,465	48,437
Treated observations	134	454	134	454	134	454
Adjusted R^2	0.310	0.160	0.202	0.118	0.186	0.160
Included in all regressions:						
Pre-spell judicial history and demographic controls	Y	Y	Y	Y	Y	Y
Year fixed effects	Y	Y	Y	Y	Y	Y
Department fixed effects	Y	Y	Y	Y	Y	Y

Notes: This table reports weighted least squares estimates of the average effect of a first closed educational centre placement on the treated. Each panel restricts the control group to the indicated type of alternative measure; unless stated otherwise the treated group is identical across panels and the entropy-balancing weights are re-estimated within each panel so that the restricted control pool matches the treated covariate profile (weighting method by panel: Covariate adjustment only: Entropy; Adding the recommendation indicators: Entropy; Adding the offence families as controls: Entropy; Reweighting on the offence families as well: Entropy). Each column corresponds to one post-spell outcome and horizon; outcomes at twelve months are defined only for spells whose full one-year post-spell window is observable before the data end and the minor's eighteenth birthday, and therefore over-represent minors placed young, while outcomes up to age eighteen accumulate events over heterogeneous windows. The counterfactual mean is the weighted control mean of the outcome. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, in parentheses, are clustered at the prescriber-department level. The sample restricts the report-linked sample of the analysiss to measures with at least one recorded nature of the case, held constant across the four panels. The first three panels use entropy-balancing weights computed on the seventeen covariates and add, successively, the recommendation indicators and twelve offence-family indicators as regression controls; the fourth panel re-estimates the weights with the offence families included in the balancing set, so that treated and control minors are also matched on the nature of the case. Offence families are built from the official nature-of-case nomenclature attached to the care episode; purely civil natures are discarded. Source: SSER data [PARCOURS, 2021-2024].

G Additional Mechanism Evidence

Table G1: Custody that closes the measure it interrupts, by type of measure

Measure	Spells	Detention within twelve months	Custody closing the measure within		
			7 days	14 days	30 days
Closed educational centre	8,381	44.8%	26.8%	35.0%	48.2%
Residential placement	30,089	22.2%	16.8%	21.7%	30.9%
Reinforced centre (CER)	7,058	20.8%	11.3%	15.1%	22.6%
Judicial supervision only	39,017	17.9%	7.1%	8.7%	12.9%
Sanctions and adjustments	20,052	4.5%	6.1%	10.8%	16.5%
Other community supervision	15,860	11.7%	3.3%	6.5%	11.4%
Suspended sentence (SME)	10,470	26.1%	3.3%	4.7%	5.9%
Pre-trial supervised release	70,412	5.5%	2.5%	4.5%	8.3%
Detention	12,063	32.5%	2.0%	6.2%	14.9%
Supervised release (LS)	19,228	5.8%	1.7%	3.6%	8.1%
Article 16bis measure	22,391	9.3%	1.1%	2.0%	3.2%

Notes: This table reports, for each type of measure on the common-support sample, the share of spells followed by a detention episode within twelve months of the decision, and among those, the share for which that episode begins within seven, fourteen or thirty days of the day the measure ends. A measure that ends on the day the minor enters custody has been closed by that custody, which is what a revocation produces; a measure that ends well before is followed by an episode it did not cause. Reading the position of the episode relative to the end of the measure without this restriction would rank measures by their duration instead: a supervision order runs for over a year and is therefore still open when custody arrives, whereas a placement of four months is usually closed. The coincidence is a proxy for revocation and not a proof of it, since a measure reaching its term shortly before an unrelated episode counts the same way; the three thresholds show how far the ranking depends on that choice. Source: SSER data [GAME/IMAGES, 2012-2022].

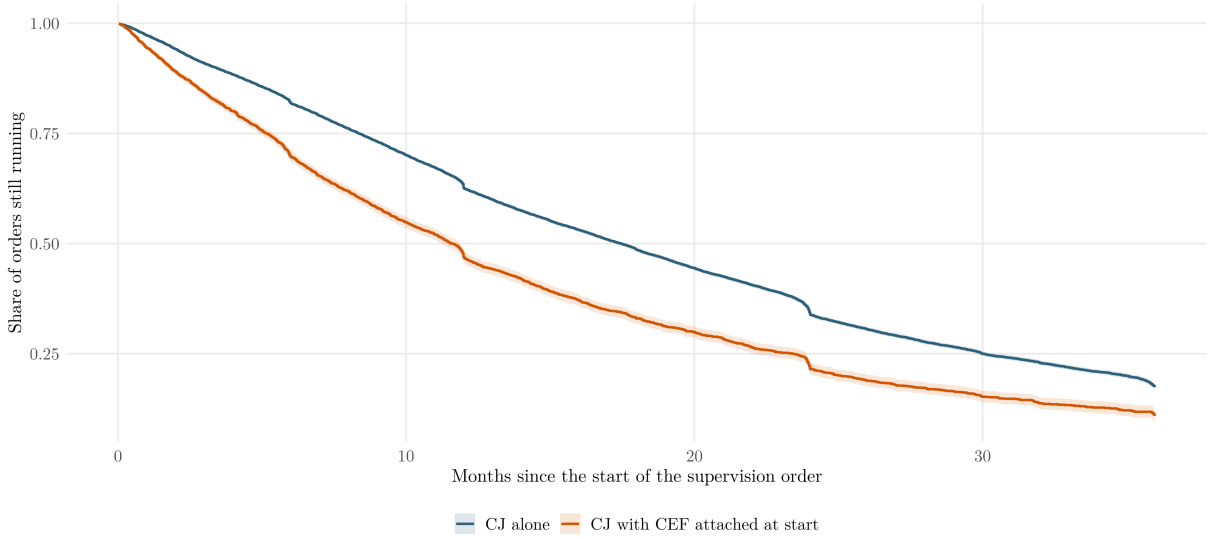

Figure G1: Survival of the judicial supervision order with and without an attached placement

Notes: This figure plots Kaplan-Meier survival curves of the duration of judicial supervision orders (CJ) started between 2012 and 2022, separately for orders with a first CEF placement attached within seven days of their start and orders never accompanied by a CEF; orders receiving a CEF later in their course are excluded to avoid immortal-time bias. Durations are censored at the eighteenth birthday, the end of the data, or an unresolved order end. Shaded areas are 95 percent confidence intervals. The comparison is descriptive: minors whose order carries a placement are not randomly selected. Median durations: 353 days for orders with an attached placement against 532 days alone. Source: SSER data [GAME/IMAGES, 2012-2022].

Table G2: Effect of a CEF placement by length of the stay, decision-anchored outcomes

	Detention episode		New CEF placement		New court referral	
	12 months	18 months	12 months	18 months	12 months	18 months
Panel A: Stay ending before four months						
Placed in a closed educational centre	0.207***	0.136***	0.039	0.030	0.021	-0.020
	(0.034)	(0.040)	(0.035)	(0.050)	(0.016)	(0.016)
Counterfactual mean	0.459	0.594	0.267	0.342	0.948	0.970
Observations	1,604	826	1,604	826	1,604	826
Treated observations	153	92	153	92	153	92
Adjusted R^2	0.240	0.338	0.203	0.276	0.122	0.200
Panel B: Stay of four months or more						
Placed in a closed educational centre	-0.062	-0.075	-0.169***	-0.175***	0.002	-0.026
	(0.047)	(0.065)	(0.038)	(0.042)	(0.022)	(0.023)
Counterfactual mean	0.403	0.528	0.245	0.314	0.952	0.975
Observations	1,584	805	1,584	805	1,584	805
Treated observations	133	71	133	71	133	71
Adjusted R^2	0.226	0.309	0.231	0.324	0.112	0.230
Included in all regressions:						
Pre-spell judicial history and demographic controls	Y	Y	Y	Y	Y	Y
Year fixed effects	Y	Y	Y	Y	Y	Y
Department fixed effects	Y	Y	Y	Y	Y	Y

Notes: This table reports weighted least squares estimates of the average effect of a first closed educational centre placement on the treated. Each panel restricts the control group to the indicated type of alternative measure; unless stated otherwise the treated group is identical across panels and the entropy-balancing weights are re-estimated within each panel so that the restricted control pool matches the treated covariate profile (weighting method by panel: Stay ending before four months: Entropy; Stay of four months or more: Entropy). Each column corresponds to one post-spell outcome and horizon; outcomes at twelve months are defined only for spells whose full one-year post-spell window is observable before the data end and the minor's eighteenth birthday, and therefore over-represent minors placed young. The counterfactual mean is the weighted control mean of the outcome. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, in parentheses, are clustered at the prescriber-department level. The sample is restricted to penal measures whose attached social-educational report recommended a closed educational centre placement. Each panel keeps the same comparison group, namely recommended minors redirected to another response, and one of the two treated groups: placements whose recorded stay ended before four months in the first panel, placements of four months or more in the second. Entropy-balancing weights are re-estimated within each panel. Outcomes run over fixed windows of twelve and eighteen months from the decision date, so that the two treated groups and their comparison group are observed over exactly the same length of time and every subsequent event counts, including a custody that revokes the placement while it is still running. The length of a stay is itself an outcome of the placement and premature departure is associated with incidents during the stay and with revocation of the order, so the two treated groups are not random sub-samples of the placed population and the contrast between panels mixes a programme-completion effect with selection on how the stay unfolded. The comparison is descriptive. Source: SSER data [PARCOURS, 2021-2024].

Table G3: Sector and programme-quality differentials in the effect of a first CEF placement

	Detention episode		New CEF placement		New court referral	
	12 months	18 months	12 months	18 months	12 months	18 months
Panel A: Sector differential						
Placed in a closed educational centre	0.161***	0.176***	0.128***	0.150***	0.137***	0.119***
	(0.008)	(0.010)	(0.013)	(0.015)	(0.010)	(0.011)
Placed × publicly run centre	0.116***	0.106***	0.006	-0.005	0.078***	0.079***
	(0.015)	(0.016)	(0.017)	(0.021)	(0.015)	(0.013)
Counterfactual mean	0.216	0.249	0.147	0.182	0.484	0.562
Observations	255,021	193,050	255,021	193,050	255,021	193,050
Treated observations	8,381	6,196	8,381	6,196	8,381	6,196
Panel B: Adding completion terciles						
Placed in a closed educational centre	0.115***	0.139***	0.100***	0.117***	0.134***	0.118***
	(0.009)	(0.011)	(0.015)	(0.017)	(0.013)	(0.013)
Placed × publicly run centre	0.058***	0.051**	-0.036*	-0.057**	0.067***	0.075***
	(0.019)	(0.021)	(0.018)	(0.024)	(0.018)	(0.018)
Placed × middle completion tercile	0.094***	0.080***	0.050***	0.068***	0.001	0.001
	(0.012)	(0.018)	(0.018)	(0.025)	(0.017)	(0.019)
Placed × lowest completion tercile	0.115***	0.104***	0.085***	0.098***	0.022	0.008
	(0.017)	(0.023)	(0.020)	(0.027)	(0.019)	(0.022)
Counterfactual mean	0.216	0.249	0.147	0.182	0.484	0.562
Observations	255,014	193,047	255,014	193,047	255,014	193,047
Treated observations	8,374	6,193	8,374	6,193	8,374	6,193
Included in all regressions:						
Pre-spell judicial history and demographic controls	Y	Y	Y	Y	Y	Y
Year fixed effects	Y	Y	Y	Y	Y	Y
Department fixed effects	Y	Y	Y	Y	Y	Y

Notes: This table reports weighted least squares estimates in which the placement indicator is interacted with characteristics of the receiving facility, within the weighted common-support sample (entropy-balancing weights aligning all eligible controls on the full treated profile, so that both sectors are compared against the same counterfactual and the same observable composition). In Panel A the baseline placement coefficient refers to accredited associative centres and the interaction tests the differential of publicly run centres. Panel B adds interactions with terciles of the facility's completion rate - the share of its terminated 2012-2022 stays reaching six months, computed on facilities with at least thirty terminated stays - with the highest-completion tercile as baseline; treated placements with an unidentified facility are excluded from the relevant panel. Facility assignment is not random: publicly placed minors carry observably heavier histories, a composition the weights hold constant but whose unobservable counterpart cannot be excluded. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, in parentheses, are clustered at the prescriber-department level. Source: SSER data [GAME/IMAGES-HEBERGEMENT, 2012-2022].

Table G4: Effect of a first CEF placement by location of the facility relative to the prescriber

	Detention episode		New CEF placement		New court referral	
	12 months	18 months	12 months	18 months	12 months	18 months
Panel A: Placed in the prescriber's department						
Placed in a closed educational centre	0.220***	0.190***	0.158***	0.182***	0.175***	0.116***
	(0.018)	(0.022)	(0.016)	(0.019)	(0.017)	(0.016)
Counterfactual mean	0.215	0.255	0.148	0.187	0.480	0.566
Observations	248,859	179,317	248,859	179,317	248,859	179,317
Treated observations	1,691	1,221	1,691	1,221	1,691	1,221
Adjusted R^2	0.186	0.217	0.102	0.126	0.178	0.192
Panel B: Placed in another department of the same interregion						
Placed in a closed educational centre	0.201***	0.224***	0.110***	0.127***	0.160***	0.154***
	(0.015)	(0.016)	(0.017)	(0.021)	(0.014)	(0.014)
Counterfactual mean	0.218	0.252	0.147	0.184	0.485	0.565
Observations	244,400	181,226	244,400	181,226	244,400	181,226
Treated observations	3,753	2,716	3,753	2,716	3,753	2,716
Adjusted R^2	0.190	0.221	0.097	0.121	0.179	0.194
Panel C: Placed in another interregion						
Placed in a closed educational centre	0.200***	0.221***	0.140***	0.152***	0.168***	0.158***
	(0.012)	(0.014)	(0.014)	(0.018)	(0.011)	(0.010)
Counterfactual mean	0.219	0.250	0.149	0.183	0.489	0.565
Observations	244,415	192,872	244,415	192,872	244,415	192,872
Treated observations	2,894	2,221	2,894	2,221	2,894	2,221
Adjusted R^2	0.200	0.229	0.098	0.124	0.188	0.204
Included in all regressions:						
Pre-spell judicial history and demographic controls	Y	Y	Y	Y	Y	Y
Year fixed effects	Y	Y	Y	Y	Y	Y
Department fixed effects	Y	Y	Y	Y	Y	Y

Notes: This table reports weighted least squares estimates of the average effect of a first closed educational centre placement on the treated. Each panel restricts the control group to the indicated type of alternative measure; unless stated otherwise the treated group is identical across panels and the entropy-balancing weights are re-estimated within each panel so that the restricted control pool matches the treated covariate profile (weighting method by panel: Placed in the prescriber's department: Entropy; Placed in another department of the same interregion: Entropy; Placed in another interregion: Entropy). Each column corresponds to one post-spell outcome and horizon; outcomes at twelve months are defined only for spells whose full one-year post-spell window is observable before the data end and the minor's eighteenth birthday, and therefore over-represent minors placed young. The counterfactual mean is the weighted control mean of the outcome. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, in parentheses, are clustered at the prescriber-department level. Each panel restricts the treated group to first placements executed in the indicated location relative to the prescriber's department, based on the department encoded in the facility unit code; the control group is the full eligible pool. The propensity score, the support and the weights are re-estimated within each comparison. The facility location is chosen by the institution and may respond to the minor's profile, so differences across panels combine placement-distance channels with residual selection into location. Source: SSER data [GAME/IMAGES, 2012-2022].

Chapter 3

The Impact of Lifting Barriers to Mental Health Care for Refugees

Evidence from a French Randomised Controlled Trial

This chapter is co-authored with Flore Gubert, Marie-Caroline Saglio-Yatzimirsky, Hala Tobji and Jérôme Valette.

Abstract

Take-up of welfare-enhancing services is often low, even among eligible individuals. Light-touch interventions have proved effective when the binding barriers are informational or administrative but whether they work when barriers are also socio-cultural, rooted in stigma and mistrust remains largely unknown. To investigate this, we study the impact of interventions aimed at overcoming barriers that prevent refugees from accessing mental healthcare, which is widely recognised as a key determinant of their integration in host countries. Building on a large governmental integration programme which supports refugees in securing employment and housing, we randomly assign 1,135 recognised refugees to a control group receiving translated materials, a confidential discussion with a trained peer-helper targeting stigma, a multilingual helpline that books mental-health appointments on refugees' behalf, or both. We find that the combined arm more than doubles the probability that we record any contact with health and mental healthcare by 7.5 percentage points on a 6.2% control mean. The helpline also raises calls but not overall engagement, and the peer discussion alone has no detectable effect. We detect no effects on employment and housing at this stage.

JEL codes: I12, I13, I14, I18, J15, J61, Z13
Keywords: Refugee integration, Mental health, Employment, Housing, Stigma

We gratefully acknowledge the members of France Terre d'Asile and Coallia for the information they generously provided to our team during this project, as well as for their time and investment. We also thank Thierry Baubet for reviewing the questionnaire and helping us adapt it to the refugee population of the AGIR programme. We are deeply indebted to Bleuène Collet, Farahnaz Faiz, Muskan Naveed, Zulfiyya Mehdiyeva and Paolo Barletta for excellent project management, as well as Mathuniga Suresh, Hakima Alizada, Hafiz Miakhel, Rohullah Sidiqullah, Ali-khan Ahmedzai, Mouni Benzakour, Mohammed Tuhin, Uhlaching Marma, Hakima Alizada, Chahrazad Boukraa, Tugay Karakuzu, Khushbo Azizi, Marwa Ahmadzai and Sibghatullah Makhdum for exceptional intervention implementation. We are greatful to Chakib Bouriche and Rayan Kaci for great data collection. This trial has been registered at the AEA RCT Registry (AEARCTR-0016069), and received IRB approval from Paris School of Economics on 21 February 2025 (2025-009). Funding for this project was provided by the IKEA Foundation, awarded by the Abdul Latif Jameel Poverty Action Lab through the Innovation for Poverty Action and Abdul Latif Jameel Poverty Action Lab Displaced Livelihoods Initiative (Grant DLI-24-02853) received on 18 November 2024. The monitoring data used in this project were provided by France Terre d'Asile and Coallia. All errors remain our own.

Contents

1 Introduction . . . **171**
2 Institutional Context and Interventions . . . **175**
2.1 Refugees in France and the AGIR programme . . . 175
2.2 The Interventions . . . 177
2.3 Theory of Change . . . 179
3 Design of the Experiment and Data . . . **180**
3.1 Sample Construction . . . 181
3.2 Experimental Protocol . . . 183
3.3 Data sources . . . 185
3.4 Variables of interest . . . 187
4 Sample description . . . **189**
4.1 Refugees' Health Profile . . . 189
4.2 Refugees' Economic and Housing Integration . . . 192
5 Empirical Strategy . . . **193**
5.1 Balance between Groups . . . 193
5.2 Intention-to-Treat estimates . . . 193
5.3 2SLS estimates . . . 195
5.4 Heterogeneous Effects . . . 195
6 Results . . . **196**
6.1 Results on take-up, engagement with mental healthcare and hospital visits . . . 196
6.2 Results on housing, employment and exit of the programme . . . 204
7 Heterogeneous effects . . . **208**
8 Conclusion . . . **210**
References . . . **212**
Appendix . . . **217**
A Implementation of the Intervention . . . **217**
A.1 Intervention materials . . . 217
A.2 Sample construction and selection into the study . . . 221
B Data and Variable Construction . . . **224**
B.1 Outcomes . . . 224
B.2 Control Variables . . . 229
B.3 Baseline Mental-Health Instruments . . . 231
B.4 Baseline Mental Health . . . 234
C Balance and Baseline Characteristics . . . **238**
D Main Estimates by Outcome Family . . . **243**
D.1 Healthcare utilisation . . . 243
D.2 Exit from the AGIR programme . . . 247
D.3 Employment and housing trajectories . . . 248
D.4 Robustness . . . 251
D.5 Effects on the Treated . . . 256
E Heterogeneous Effects . . . **261**
E.1 Effects by Gender . . . 261
E.2 Counterfactual Peer Assignment . . . 269

1 Introduction

Many welfare-enhancing programmes and services suffer from persistently low take-up, even among individuals who are formally eligible and would clearly benefit from them (Currie, 2004; Ko and Moffitt, 2022). A growing literature in economics shows that light-touch, relatively inexpensive interventions can significantly increase take-up when the binding barriers are primarily informational or administrative in nature (see, among others, Bhargava and Manoli, 2015a; Domurat et al., 2021; John and Orkin, 2022). Whether this logic extends to settings where barriers are not purely informational, but also deeply socio-cultural (involving stigma or cultural dissonance) is far less understood, and considerably more demanding to establish.

Refugee mental healthcare offers a particularly stringent test case for this question. According to the latest UNHCR annual report, the number of forcibly displaced individuals reached record levels in 2025, driven by escalating conflicts, human rights violations, and natural disasters (UNHCR, 2025). In Europe, the refugee population increased from 7 million in 2021 to 13.6 million in 2025. In France, applications for international protection rose from 104,577 in 2021 to 153,715 in 2024 (OFPRA, 2025). These large inflows have generated significant integration challenges for host countries, as refugees consistently exhibit lower employment rates and poorer housing conditions than both other immigrant groups and the native population (see, among others, Dustmann et al., 2017; Brell et al., 2020; Fasani et al., 2021a).

Among the various factors influencing refugees' economic integration in their host countries, observational studies suggest poor mental health as a significant barrier to both immigrant and refugee integration (Chatterji et al., 2007; Ojeda et al., 2010; Banerjee et al., 2017; Ruiz and Vargas-Silva, 2018; Di Thiene et al., 2021; Niederkrotenthaler et al., 2020; Dietrich et al., 2023; Dang et al., 2023; Foged et al., 2026).[1] Indeed, refugees represent a highly vulnerable population with a high risk of mental disorders, stemming from both the challenges of integration and the enduring effects of violent or life-threatening experiences in their countries of origin or during displacement.[2] The high prevalence of mental disorders among refugees is closely linked to the various barriers that limit their access to mental health care in host countries.[3] On the supply side, existing challenges relate to legal access to care (Fundamental Rights, 2011) and the capacity of health systems to provide adequate services to refugees (Norredam et al., 2005). On the demand side, two types of barriers are typically distinguished (Satinsky et al., 2019; Byrow et al., 2019). On the one hand, structural barriers include the lack of information and the challenge of navigating a foreign healthcare system, which often differs significantly from that of one's home country and involves many different actors (Salami et al., 2019; DeSa et al., 2022) as well as language difficulties. On the other hand, socio-cultural barriers, such as limited

[1] See Ridley et al. (2020) for causal evidence on how mental illness treatment improves well-being and labour market outcomes for the overall population.

[2] See Steel et al. (1999), Burnett and Peel (2001), Turner et al. (2003), Derluyn et al. (2004), Barenbaum et al. (2004), Fazel et al. (2005), Johnson and Thompson (2008), Frijters et al. (2010), Teodorescu et al. (2012), Bogic et al. (2015), Langlois et al. (2016), Tinghög et al. (2017), Abel et al. (2019), Giuntella et al. (2018), Couttenier et al. (2019), Gleeson et al. (2020), Blackmore et al. (2020), Mellor et al. (2021), Stillman et al. (2022), and Bryant et al. (2023).

[3] The salience of these constraints is illustrated by Liu et al. (2025) who show that, following the 2008 Great Recession in Italy, natives increased their mental healthcare use while immigrants did not, despite the crisis.

awareness, stigma, and cultural dissonance, have been shown to be particularly salient in refugee communities (Kiselev et al., 2020). Refugee mental healthcare is therefore a setting in which structural and socio-cultural frictions are simultaneously present and particularly acute, making it an especially demanding context in which to test whether low-cost interventions can improve access. A related literature has tested whether such low-cost interventions can help refugees and immigrants overcome integration barriers in other domains (see, among others, Battisti et al., 2019; Rozo and Grossman, 2025; Batista et al., 2025). Whether this logic extends to a domain where the socio-cultural frictions identified above are so central remains, however, untested.

This paper provides causal evidence from a field experiment testing whether targeted, low-cost interventions can alleviate demand-side barriers to mental healthcare among recently recognised refugees in France. The experiment takes place within the AGIR programme, the French government's main integration policy supporting refugees' access to employment and housing, and more specifically in two departments (Paris and Essonne) where it is implemented by France Terre d'Asile and COALLIA, respectively.[4] A total of 1,403 refugees enrolled in AGIR and consenting to participate in the experiment were randomly assigned at the individual level, with stratification on *Groups of nationality* $\times$ *gender*, to one of four arms.[5] Individuals in the control group receive a multilingual information leaflet listing mental-health services tailored to refugees. The first treatment group consists in a peer-helper arm which offers, in addition, a one-on-one confidential discussion in the participant's native language with a trained peer-helper of similar migration background, designed to alleviate stigma and mistrust towards mental-health professionals. The second treatment group is a helpline arm providing access to a multilingual telephone service staffed by peer-helpers who schedule appointments on the participant's behalf at partner mental-health facilities, designed to lift structural and navigational barriers. Finally, a third treatment arm combines the two interventions, testing whether the socio-cultural and structural channels are complementary.

Our theory of change posits that relaxing the aforementioned barriers increases engagement with formal mental healthcare, which may in turn supports progress on the employment and housing outcomes, which are the core objectives of the AGIR programme. To evaluate the effects of the different interventions, we thus combine several layers of measurement from multiple sources, including (i) housing and labour market integration from AGIR implementers' monitoring data; (ii) self-reported measures of mental health and economic integration collected through baseline and endline surveys; (iii) mental healthcare take-up recorded by our four partner service providers in Paris; (iv) administrative data on healthcare utilisation from the Paris public hospital system; and (v) [plausibly additional indicators of economic integration from public administration on health and labour market outcomes].

[4]The AGIR programme (*"Accompagnement Global et Individualisé pour l'accompagnement des Réfugiés"*) was launched in 2022 and now operates nationwide. It provides a 24-month individualised support pathway focused on housing and employment. Entrants are pre-screened by the OFII (*"Office Français de l'Immigration et de l'Intégration"*) to identify particularly vulnerable individuals, notably those facing precarious housing and weak labour market attachment, precisely the population for whom mental health frictions are likely to be most binding. AGIR does not include a dedicated health component, allowing our intervention to be evaluated as a scalable module within an existing national policy rather than as a stand-alone research protocol.

[5]Recruitment took place in Paris between May and late November 2025 and in Essonne from March to late June 2026.

Preliminary results based on the Paris sample only show that the intervention generates modest but statistically significant effects on engagement with mental healthcare. In particular, the combined treatment, which includes both helpline and the peer-helper component, increases the probability that a participant exhibits at least one post-treatment signal (helpline use, contact with partner facilities providing mental health support, or hospital visit) by 7.5 percentage points relative to a control mean of 6.2%. This effect appears to be driven mostly by the female sample, and the identity of the peer-helper also seems to matter. By contrast, the peer-helper-only and helpline-only arms have no significant impact. Consistently, the helpline significantly increases take-up. Treated individuals are 3 to 4.3 percentage points more likely to call the helpline and, conditional on calling, most interactions translate into scheduled and attended appointments. These effects do not systematically translate into increased use of our partner mental health facilities; instead, we find suggestive evidence that individuals in the combined treatment arm may also be seeking care outside the programme's partner network. Given the small magnitude of the intervention's effect on healthcare use in general and mental healthcare in particular, combined with the short time elapsed since our intervention, we find no detectable impact on employment or housing status trajectories. We are nevertheless left with the puzzling finding that our intervention appears to hinder refugees' positive exit from the AGIR programme prior to the mandatory 24-month administrative cap.

This paper contributes to both academic research and policy debate by providing causal evidence on the role of mental health in refugees' integration and by evaluating cost-effective interventions to improve access to care. On the academic side, it offers, to our knowledge, one of the first randomised evaluations of interventions targeting demand-side barriers to mental healthcare among recently recognised refugees in a high-income host country. Indeed, existing experimental work on mental health among displaced populations has mainly focused on humanitarian settings in low-and middle-income countries (Islam et al., 2024; Moya et al., 2024; Smith, 2025) or on undocumented immigrants in urban US settings (Sabety et al., 2023), while the broader literature on refugee integration in OECD destinations has paid comparatively little attention to mental health as a policy-relevant determinant of integration. In addition, our four-arm experimental design allows us to disentangle two distinct demand-side channels, socio-cultural stigma and structural navigation, which have been mostly studied in isolation in previous studies. On the policy side, the intervention was designed to be directly scalable within an existing national integration programme. The peer-helper and helpline components effectively add a mental health module to AGIR, and directly address a gap identified by programme operators as key constraints on the programme's effectiveness.

This study speaks to several strands of the literature. First, this paper contributes to the literature on the role of information and navigation frictions in access to public services, particularly in healthcare. The closest paper to our analysis is Sabety et al. (2023), who show in a randomised experiment among undocumented immigrants in New York City, that providing active assistance with appointment scheduling substantially increases healthcare utilisation. Consistent with these findings, our results indicate that the helpline intervention, by offering direct assistance in navigating the healthcare system, generates larger behavioural responses than

passive information provision alone, in a context characterised by institutional complexity and immigrant-specific frictions (Liu et al., 2025). This paper also speaks to the literature on stigma and take-up on welfare and health programmes (Henderson and Thornicroft, 2009; Bhargava and Manoli, 2015b; Celhay et al., 2025; Yu, 2023; Yang et al., 2023). By testing a structured, peer-mediated intervention explicitly designed to counter stigma, we provide direct evidence on whether this channel can be addressed at scale, and, via the contrast with the helpline arm, whether a brief peer-mediated interaction is sufficient to do so. In contrast to Smith (2025), who find that monetary incentives increase the diffusion of a mental health awareness campaign among Syrian refugees in Jordan, highlighting the role of stigma-related barriers, we find no evidence that peer-based interactions alone increase engagement with mental health services. This suggests that stigma may be particularly persistent in this context, and that reducing organisational barriers may be a more effective lever for increasing take-up.

Second, this paper contributes to the limited body of work on the impact of providing mental health support to displaced communities. In the development context, Islam et al. (2024) randomly assigns 3,500 Rohingya mothers in Bangladeshi refugee camps to receive psycho-education and parenting guidance. The intervention leads to a significant reduction in psychological trauma and depression among both mothers and their children.[6] Similarly, Moya et al. (2024) evaluate the impact of providing mental health support to forcibly displaced caregivers in Colombia, and find significant improvements in caregivers' mental health eight months after the programme. Our study complements this work by shifting the focus to recognised refugees in a high-income host-country where stable and formal healthcare exists, but where navigation and cultural barriers bind on the demand side.

Finally, this paper also adds to the broad and growing literature on the economic integration of refugees (Dustmann and Görlach, 2016; Dustmann et al., 2017; Fouka, 2023; Bahar et al., 2024; Foged et al., 2024a). This literature has unveiled several factors or policies that shape refugees' integration, among which the level of violence in origin countries (Aksoy et al., 2024; Couttenier et al., 2019), initial conditions upon arrival (Braun and Dwenger, 2019; Abramitzky et al., 2023), intended duration of stay (Adda et al., 2022), language barriers and training (Lochmann et al., 2019; Foged et al., 2024b), financial assistance (Rosholm and Vejlin, 2010; Dustmann et al., 2024), legal status (Fasani et al., 2021b; Urbina et al., 2023), or labour and job search assistance (Battisti et al., 2019; Andersson Joona et al., 2016). However, comparatively, little attention has been paid to mental health as a determinant despite its prominence in observational work. Exceptions exist with Dang et al. (2023) exploiting variations in refugees' traumatic experiences in Australia, interacted with post-resettlement periods, to causally demonstrate that poorer mental health significantly lowers the likelihood of employment and reduces labour income. Similarly, Foged et al. (2026) find that nearly one-third of 2022 arriving Ukrainian refugees in Denmark suffer from Post-Traumatic Stress Disorder (PTSD) or

[6]Some publications involving the Bangladesh-based organisation GDRI have been subject to scrutiny in the context of replication efforts (see Brodeur et al. (2025) and Brouwer et al. (2023) and Brodeur (2025) for further discussion). As of today, these concerns do not pertain specifically to the present study but highlight the general importance of transparency and replication in this field. For additional context, see the *AEJ:AE*'s expression of concern, and Retraction Watch's report.

Complex-PTSD (CPTSD), which lowers their employment probability by 6 percentage points and accounts for one-fifth of the refugee-native employment gap after 18 months. They show that this gap is unlikely to be driven by selection on unobservables. This paper adds to this literature by identifying access to mental healthcare as a policy-relevant and causally identified lever for improving integration outcomes. In doing so, it complements the employment- and housing-focused instruments that existing integration programmes typically deploy.

The remainder of this paper is organised as follows. Section 2 provides background information on the context in which the programme took place and describes the intervention and its objectives, as well as the experimental design. Section 3 describes our sampling frame, the data sources and the construction of the main variables. Section 4.2 details the mental health and economic profile of the participants. Section 5 presents the balancing tests on participant's baseline characteristics and the estimation strategy. Sections 6 and 7 present the results and heterogeneity analyses. We conclude in Section 8 by discussing implications.

2 Institutional Context and Interventions

This section presents the main features of refugee integration policy in France, with a particular focus on the AGIR programme. We then describe the interventions and their implementation. Finally, we outline the theory of change underlying our analysis.

2.1 Refugees in France and the AGIR programme

France has a long-standing tradition of providing protection and asylum to displaced populations, shaped by its political and humanitarian commitments. However, in recent years, migration debates have intensified and become increasingly polarised, with policymakers navigating the balance between upholding humanitarian obligations and responding to growing political pressures to restrict inflows.

Since 2015, the country has seen a sharp rise in asylum applications, peaking at over 105,000 submissions to the French Office for the Protection of Refugees and Stateless Persons (OFPRA hereafter) in 2019. Although applications decreased to around 75,000 annually in the following two years, they have increased since then (OFPRA, 2024). In 2025, OFPRA received approximately 145,210 applications for international protection. The top five countries of origin for asylum seekers were Afghanistan (13,800), the Democratic Republic of the Congo (13,240), Haiti (12,600), Ukraine (12,310), and Guinea (8,640), followed by Côte d'Ivoire (7,530) and Turkey (6,830). (OFPRA, 2025). The overall protection rate for the same year was 41.2%.

Beyond the numbers, concerns about refugee hosting in France have largely focused on the low level of economic assimilation among refugees and the state's capacity to absorb and integrate such a vulnerable population. A representative survey conducted by the Ministry of the Interior tracked newly arrived migrants over the four years following the issuance of their first residence permit in 2018 (ELIPA2, 2024b). Among these migrants, 17% were recognised refugees

or internationally protected persons[7]. Although the survey pointed out notable improvements in their situation during this period, by 2022, only 62% of refugees were employed and, respectively 6% and 7% remained in collective accommodation centres or makeshift housing compared to 15% and 12% in 2019, respectively (ELIPA2, 2024c). Importantly, 47% reported having mental health issues one year after receiving their residence permit, declining to 30% four years later.[8]

In response to these concerns, the government has reinforced its asylum system by expanding reception capacities and integration programmes while simultaneously tightening immigration policies. In 2022, the French government launched the AGIR programme (*Accompagnement Global et Individualisé des Réfugiés*), which has been gradually implemented across departments (see Figure A1 in Appendix). In each department, an AGIR operator (typically a non-profit organisation or a consortium of such organisations) is designated by the Prefect following a public procurement procedure, and becomes the single point of contact for the integration of recognised refugees. Concretely, the AGIR programme offers a 24-month support pathway with a focus on housing and employment to individuals recognised as beneficiaries of international protection (including both refugees and beneficiaries of subsidiary protection) who obtained this status in the current year or the preceding year, and have signed the Republican Integration Contract (*Contrat d'Intégration Républicaine*, CIR) with the French Office for Immigration and Integration (*Office Français de l'Immigration et de l'Intégration*, OFII).[9] Once they enter the programme, refugees are followed up by a pair of "integration advisors" (one specialised in access to rights and housing, and the other in employment and vocational training) who commit to regular meetings at least every two months to review their progress and provide tailored guidance at each stage of their integration pathway.[10]

However, not all refugees with genuine support needs in these areas are referred to the AGIR programme, and the criteria applied by OFII in the selection process are found to vary depending on the extent to which the active caseload initially set at the departmental level is reached. What is clear, however, is that the OFII refers only the most vulnerable refugees to the AGIR programme, a tendency that has become even more pronounced since July 2024, when,

[7]The term used in ELIPA2 is "humanitarian reasons"; we will use this term when referring to the survey throughout the rest of the paper.

[8]A person is considered having mental health issues if they respond yes to one of these questions: "Over the past 12 months, have you ever felt particularly sad, down, or depressed, for most of the day, nearly every day, for a period of at least 2 weeks?" and "Over the past 12 months, have you ever felt, for a period of at least two weeks, that you had lost interest in everything, or lost interest or pleasure in the things you usually enjoyed?". We note that our baseline survey includes two similar questions, but asks only about the past two weeks rather than the past 12 months.

[9]The CIR is part of the personalised integration pathway designed for non-European foreign nationals admitted to reside in France who wish to settle there on a long-term basis. It is concluded between the French State and the signatories, who commit to completing the training courses prescribed to them in order to facilitate their integration into French society. The CIR includes language training and civic education, which enable foreign nationals to meet the language proficiency requirements and pass the civic examination that have been mandatory since January 1, 2026 for those seeking to obtain a multi-year residence permit or a permanent residence card (translated from https://www.service-public.gouv.fr/particuliers/vosdroits/F17048. Accessed: April 21, 2026)

[10]The cost of the AGIR programme is difficult to assess, as budget allocations are not reported as a stand-alone item. The only executed figure available from a primary source, the *Cour des comptes*, indicates that €16 million were actually spent, compared to an initial forecast of €76 million. A multi-annual envelope of €630 million for the period 2023–2027 is cited by several sources, but we have not been able to independently verify this figure in an official report.

driven by budgetary pressures, referrals were abruptly curtailed or even frozen in a number of departments (OFII, 2024).[11] This is reflected in the profile of programme entrants: 78% are neither employed nor enrolled in any form of vocational training, and respectively 37%, 28% and 8% are living in refugee centres, staying with a relative or homeless (DSED, 2026a). As of December 31, 2025, a total of 23,100 refugees had completed their support pathway and exited the programme, of whom 14,400 did so in 2025 alone (DSED, 2026b). For 37% of these beneficiaries, the exit was classified as positive, meaning that the support provided in the areas of employment and/or housing concluded with a stable and sustainable outcome. More specifically, 64% of those who received housing support exited the programme with stable housing conditions, while the positive exit rate for employment stood at 42%. Encouraging as these results may be, they nonetheless fall short of the 60% positive exit rate target set for the 2023-2025 period. This is partly attributable to structural features of local labour markets and housing stock, but also to other barriers, notably the mental health of refugees. While frequently highlighted by AGIR operators as a critical unmet need among their beneficiaries, mental health support falls outside the explicit scope of the programme. It is precisely against this backdrop that our interventions were designed.

2.2 The Interventions

The experiment was carried out in the Paris and Essonne departments, where the AGIR programme is implemented, respectively, by France Terre d'Asile (FTDA, hereafter) and COALLIA.[12] The former is a solidarity organisation founded in 1971, whose primary mission is to support asylum seekers and defend the right to asylum in France. The latter is a non-profit organisation founded in 1962 that provides housing, accommodation, and social support services to vulnerable populations.

Both associations granted us access to their premises and facilitated outreach to refugees enrolled in their programme. This included individuals already registered at the start of the experiment, as well as those who joined over the course of its implementation. This close collaboration allowed us to access integration tracking data recorded through the AGIR information system and helped us minimise attrition by benefiting from intensive monitoring of each refugee in the programme. The choice of these two departments, driven primarily by logistical considerations, implies that the eligible population may differ from the broader refugee population enrolled in the AGIR programme in France, as it includes only those located in Paris and one of its inner suburbs, rather than in other French departments.

The study randomly assigns refugees enrolled in the AGIR programme in Paris and Essonne to one of four groups (three treatment arms and one control group), that are described below.

(C) Control Group: Basic information. Participants receive a standardised information

[11] A Senate information report filed on February 11, 2025 specifies that the programme aims to reach an average of 25,000 beneficiaries nationwide (compared to 31,446 beneficiaries as of July 31, 2024), out of approximately 50,000 eligible refugees per year (Information Report No. 326, 2024–2025)

[12] FTDA and COALLIA are also the operators implementing the programme in, respectively, 6 and 18 other French departments.

guide in their language[13] listing a selection of available mental health services in the Paris region.[14] It includes information on the name, address, nearest metro station, opening hours, and method of consultation (with or without appointment, or both), as well as telephone numbers and email addresses. A "RISE Partner" stamp is also added to the leaflet, allowing refugees to easily identify services that are affiliated with the project.[15] Figure A3 depicts the English version of the information leaflet.

(G1) Treatment 1: Basic information and discussion with a peer-helper. In addition to the standard information guide (C), participants have a one-on-one confidential discussion with a peer-helper, in their language. The great majority of peer-helpers are themselves refugees with a similar background and lived experience, recruited through a university diploma programme in mediation and interpretation in a migration context offered by INALCO (Paris).[16] The interview focuses only on mental health issues and mental health care structures tailored to refugees in Paris. Project staff do not have access to this discussion or its content. As supported by the literature review, this treatment conjectures that face-to-face interactions between refugees and trained peer-supporters can reduce sociocultural barriers, self-stigma, and mistrust towards mental health professionals.

(G2) Treatment 2: Basic information and access to a helpline. In addition to the standard information guide (C), participants are granted access to a helpline that provides personalised support from a peer-helper who speaks their language[17], specifically to assist with booking appointments at mental health facilities. The phone number for this helpline is provided by attaching a label to the back of the information leaflet. There is one telephone number per refugee language. Figure A4 depicts this label. The sticker is placed in front of the refugee when the information guide is handed over to increase the salience of this treatment. The same individuals who provide peer support are responsible for answering the helpline and scheduling appointments at a facility for the refugee who calls, using a phone provided to them. It should also be noted that this assistance is limited to scheduling the first appointment only. As supported by the literature review, this treatment conjectures that providing telephone assistance leading to appointments at appropriate care facilities can eliminate structural barriers to accessing mental health care.

(G3) Treatment 3: Basic information, discussion with a peer-helper and access to a

[13]The languages available are French, English, Arabic, Pashto, Dari, Urdu, Bengali, and Tamil, which account for the majority of languages spoken by refugees in AGIR Paris. Turkish was added for AGIR Essonne, where it is the most widely spoken language among refugees. In the remainder of the paper, we use the terms "preferred language" or "their language" for ease of reading.

[14]The list includes associative structures (*Le Chêne et l'Hibiscus, COMEDE, Centre Primo Levi, Centre Minkowska and Parcours d'Exil*) and the *Centre Régional du Psychotrauma Paris Nord* of Avicenne Hospital. All cater for a large proportion of either asylum-seekers or refugees and offer free transcultural psychological support by specialised psychotherapists with the support of interpreter-mediators.

[15]The WATIZAT leaflet inspires our information leaflet, but the latter only includes information on mental healthcare services. See https://watizat.org/en/accueil/leaflets-paris/ and https://orspere-samdarra.com/wp-content/uploads/2021/12/os-outil-sante-mentale-fr.pdf.

[16]See https://www.inalco.fr/formations/du-hospitalite-mediations-migrations-h2m for additional information on this diploma.

[17]The same languages are available as those described for the information guide.

helpline. In addition to the standard information guide (C), participants receive both the peer-helper discussion (G1) and helpline access (G2). This treatment is thus a combination of the two previous treatments and is based on the hypothesis that these two approaches are complementary. By simultaneously addressing both sociocultural and structural barriers to accessing care, it aims to improve the ability of refugees to seek treatment.

2.3 Theory of Change

Figure 3.1 presents the study's theory of change. First, it posits that face-to-face discussions between refugees and trained peer-helpers can address sociocultural barriers to mental healthcare (**Assumption 1**). By communicating as equals in the same language and with similar backgrounds, peer-helpers can reduce stigma, increase mental health awareness, and reduce distrust toward mental health professionals. This intuition is supported by a randomised intervention with 849 Syrian refugee friend groups in Jordan, which finds that the disclosure of monetary incentives increase the WhatsApp sharing of a mental health awareness campaign offering free counselling by 34% (Smith, 2025). These results suggest that payments act as a "stigma-alleviating excuse" for senders, therefore highlighting significant stigma-related barriers to mental health support among refugee communities.

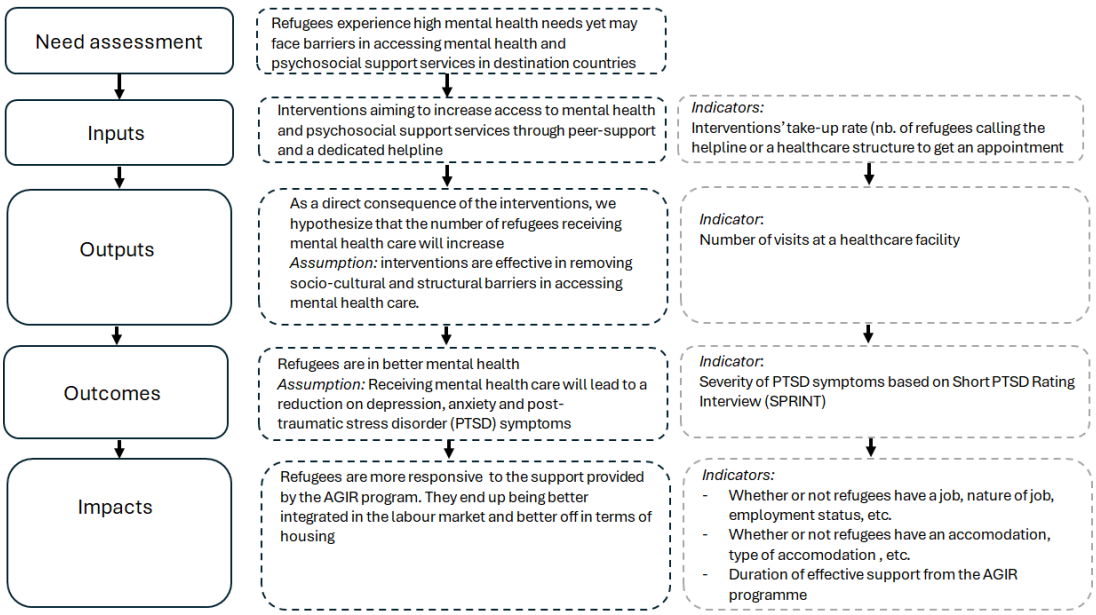

Figure 3.1: Theory of change

Additionally, our theory of change rests on the assumption that offering a helpline for appointment scheduling, simplifying healthcare navigation, and overcoming language barriers remove structural obstacles and increase refugees' use of formal mental health services (**Assumption 2**).[18] Close to our design, Sabety et al. (2023) conduct a randomised experiment focusing on undocumented immigrants in New York city, in which participants in the

[18]It is worth noting that since all interactions with refugees are conducted in their native language (even for the control group), we are not be able to isolate the specific impact of language support.

treatment group received active assistance in scheduling an appointment at one of nine partner clinics, while participants in the control group received only information about available support services but no active assistance in accessing them. Their results show that after three months, 57% of the treated individuals visited the partner clinics, compared to only 16% in the control group. Emergency department visits among treated individuals also decreased substantially and significantly compared to the control group, suggesting an improvement in their health conditions.

Ultimately, our experimental design evaluates the impact of these two types of intervention, and their joint effect, on mental healthcare utilisation and status, using self-reports from baseline and endline surveys, as well as administrative data provided by partner mental health support structures in Paris. Our theory of change posits that improving access to mental health support for refugees improves their mental health and, in turn, facilitates economic integration (**Assumption 3**). The latter is measured through employment and housing outcomes reported by the AGIR programme, as well as data collected in the endline survey and other administrative databases.

Compared to Smith (2025), our objective is not to reveal the existence of social stigma in sharing information about mental health support. Instead, our project aims to lift these barriers, thereby increasing mental healthcare take-up and improving economic integration. Thus, our study largely benefits from being embedded within a large-scale individualised support programme aimed at improving refugees' integration. This framework allows us to track integration outcomes from the moment refugee status is granted and for several months after the programme's ending. Importantly, partnerships with medical institutions allow us to measure the take-up of mental health support among refugees. Our experimental design thus provides an ideal setting for maximising the likelihood of increased uptake following interventions. Moreover, in our context, peer-helpers are trained professionals who have received standardised training in social mediation. Unlike informal peer networks, they are compensated for their work and have no prior personal relationships with the refugees they assist. This design eliminates the social stigma costs associated with providing information on mental health support, thereby maximising the likelihood of refugees engaging with mental health support. Finally, our helpline is designed not only to provide information but also to actively assist refugees in securing an appointment by scheduling one on their behalf at an available healthcare facility.

3 Design of the Experiment and Data

This section first presents the experimental protocol, including the sampling frame, the recruitment procedure, and the randomisation process. It then describes the data sources and variables of interest. Figure 3.2 summarises the flow of refugees from the sampling frame to the final sample. As we were granted formal access to FTDA and COALLIA's monitoring systems, we are able to characterise the population constituting our randomisation sample, the individuals we were able to approach, and those who consented versus those who did not. This level of detail is rare in this type of RCT and worth documenting, each stage is therefore described below, along

with descriptive statistics characterising the population at each step.

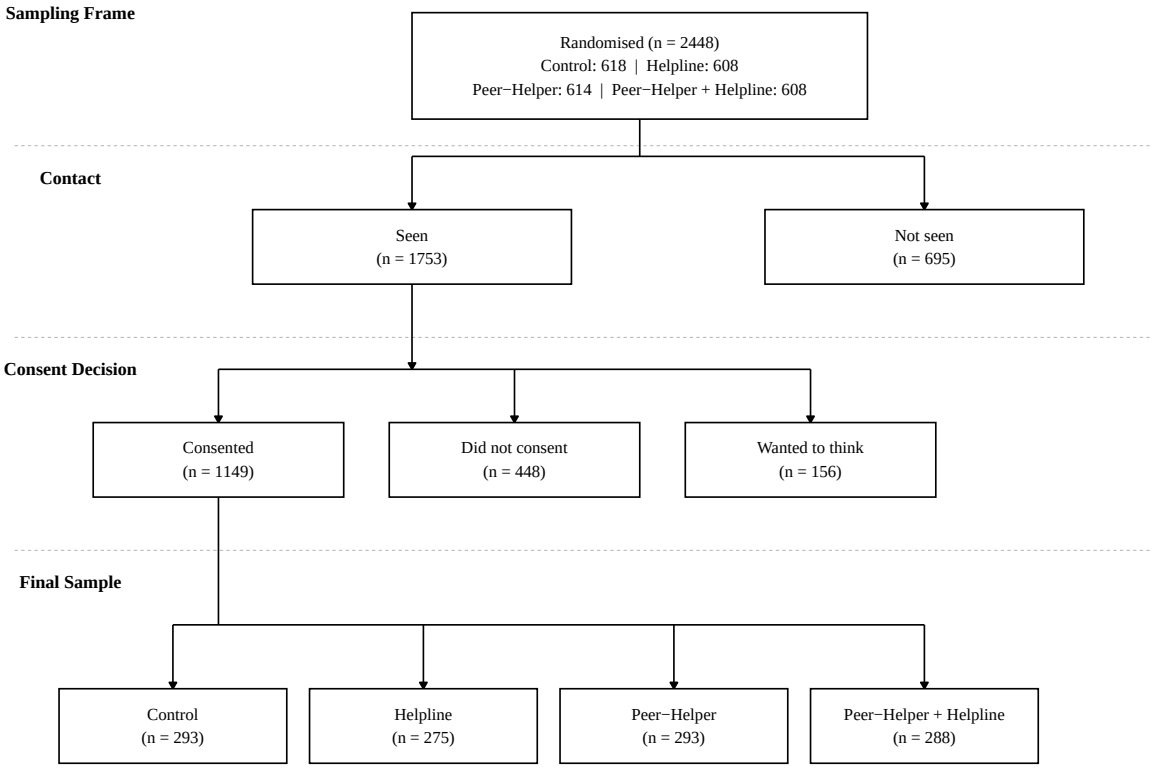

Figure 3.2: Flow diagram of participants (FTDA sample)

Notes: Individuals seen by RISE are those who met with the project managers and received the project explanation. Among those, some consented, some did not consent, and some requested additional time to decide ("Wanted to think"). For the latter group, we were unable to collect a final decision before the end of the inclusion period. The bottom four boxes show the final distribution of participants across the four treatment arms. This figure concerns FTDA only, as COALLIA data collection is still ongoing.

3.1 Sample Construction

Sampling Frame. The list of refugees actively enrolled by FTDA in Paris and COALLIA in Essonne at the time of our experiment's launch and throughout its duration constitutes the universe of the population eligible for our interventions.[19] As of May 12, 2025, the day of randomisation for FTDA Paris refugees, 2,097 individuals were actively enrolled. As of March 23, 2026, the corresponding figure for COALLIA was 598. An additional 337 and 98 refugees were randomised in subsequent waves on September 15 and November 10, respectively, as new participants entered the AGIR programme in Paris over the course of the experiment.[20]

[19]The active phase of the experiment in Paris started on May 26, 2025 and ended on November 28, 2025. In Essonne, the experiment started on March 30, 2026 and [is scheduled to end in late June 2026].

[20]The sum of the three randomisation waves does not match the total number of randomised individuals reported in Figure 3.2. This discrepancy arises from the pseudonymisation procedure applied prior to randomisation,

In May 2025, at the beginning of our experiment, the average participant assisted by FTDA in AGIR Paris was a male (78%) aged 32 (between 18 and 77 years). Five nationalities accounted for 68% of beneficiaries: Afghanistan (33%), Bangladesh (19%), Côte d'Ivoire (7%), Sri Lanka (6%), and Somalia (3%). 92% reported needs related to employment, housing, and support with administrative or legal rights at entry. Among the new entrants, 64% lived in an accommodation provided by a third party, 12% were either homeless or lived in precarious conditions, and 19% were housed in public assistance shelters. Furthermore, 46% of participants entered the programme unemployed.

RISE Contact. Figure 3.2 shows that 2,448 individuals were enrolled in the AGIR FTDA programme at the time of the experiment and were randomised, of whom 1,753 were informed of the project and invited to participate.[21] We are able to conduct balance checks on some characteristics between individuals seen versus not seen by the project managers. Figures A6 shows this comparison. Baseline characteristics appear broadly balanced. However, women were less likely to be seen by the project. This is unsurprising, as anecdotal evidence from the field suggests that women enrolled in AGIR attend meetings with social workers less frequently, with their husbands or children sometimes attending in their place. Additionally, some attended with young children[22] and may not have had the time or energy to engage with the project presentation. These findings suggest that new strategies may be worth considering to improve outreach to women.

Consent Decision. After meeting with the project managers and receiving the project explanation, refugees could choose to consent, decline, or request additional time to decide. Figure A7 shows a broadly stable weekly consent rate of approximately 66%, and Figure A8 shows the cumulative inclusion of refugees in RISE broken down by consent status. Figure A9 reports balance checks between those who consented and those who did not, and baseline characteristics appear broadly balanced across the two groups. As with the seen versus not seen comparison, women tended to be more hesitant and more likely to consult their partners before making a decision. We also note that more educated individuals display higher consent rates, possibly because they were more familiar with the concept of a research project.

Final Sample. The final sample of refugees participating in our study differs from the initial lists of refugees enrolled in the AGIR programme maintained by FTDA and COALLIA, as participation requires both meeting with each refugee to present the project and obtaining their informed consent.

which is anchored to a combination of identifying variables (birth name, first name, date of birth, nationality, and AGFREF number). When any of these variables changed for a given individual, a new pseudo-identifier was generated, resulting in duplicate entries. All such cases were identified and resolved by assigning each individual to the treatment arm of their first randomisation.

[21] Some individuals were not reached for various reasons, most commonly because they rarely met with their AGIR social workers.

[22] Although AGIR Paris has an on-site childcare facility, it was not always available.

3.2 Experimental Protocol

We now describe the procedure through which refugees were approached, informed about the project, and enrolled. During their scheduled individual meetings with refugees, caseworkers briefly introduced the RISE project (Refugees' Integration Through Support and Mental Health Enhancement) using a scripted presentation designed to ensure consistency across interactions. This presentation was intentionally brief to preserve the trust established between caseworkers and refugees, especially given the stigma surrounding mental health issues among our population of interest. In addition, we wanted to minimise the risk of confusion between the support offered under the AGIR programme and the RISE project. For these reasons, caseworkers simply invited refugees to meet with a member of our team for further information, which also served to limit their involvement in the project and prevent any interference with the interventions.[23]

Once a refugee was referred to our team, our project managers showed him/her a brief introductory video about the RISE project in the language of his or her choice, in the same set of languages as the information guide, with the Arabic version available in both Standard and Algerian Arabic. The English version of the video script is available in the appendix (see Figure A5).[24] After the video was shown, a consent form translated into the same languages listed above was provided to ensure full understanding of the project's objectives, the nature of the intervention and data collection, and the refugee's rights regarding data use and the possibility to withdraw from the project at any time.[25] Accommodations were made for illiterate refugees, who had the consent form read aloud to them. Individuals were enrolled in the project, and the intervention began only after they had signed the consent form. Of the initial 1,149 individuals (see Figure 3.2 for the exact distribution among treatment groups), fourteen of them had already left the AGIR programme by the time their treatment date came up, and are dropped from every estimation because the outcomes we measure are all conditional on being supported by the programme. The estimation sample is therefore 1,135, and the exclusions are spread evenly across arms: three each in the control, helpline and peer-helper arms, five in the combined arm. These figures cover the Paris site. In Essonne, where AGIR is run by COALLIA, 255 refugees consented, of whom 23 had left the programme before their treatment date, leaving 232 in four balanced arms of 66, 61, 66 and 62. Treatment there ran from 30 March to 30 June 2026, so at the July 2026 extraction the median refugee had been followed for under two months and the last ones for a few weeks. That is too short for outcomes that are counts of events over a follow-up window, and far too short for exit from a programme capped at twenty-four months. We therefore restrict every estimate reported below to Paris, and will incorporate Essonne once its follow-up is comparable.

Assignment to Treatment. Individuals are assigned to one of the three treatment arms or the control group through a stratified randomisation procedure at the individual level. In so doing,

[23] To further ensure that refugees clearly distinguish the RISE project from the AGIR programme, we developed a separate visual identity. This includes the logo shown in Figure A2, along with a distinct colour scheme, posters, and other supporting materials.

[24] For refugees whose native language was not spoken by any peer-helper, *e.g.*, Ukrainian refugees, we assigned a peer-helper speaking one of the refugee's other spoken languages, if any.

[25] The English version of the consent form is available upon request to the authors.

we create an exogenous source of variation in access to information and support regarding mental health, orthogonal to both observed and unobserved characteristics of the refugee population in the AGIR programme. As a result, any differences in mental health access, employment, or housing outcomes between groups are causally attributable to the different components of our intervention.

For participants already enrolled in AGIR, randomisation is carried out once, using the list of AGIR participants as of the project's start date. This list is extracted from ANITO and SILAO, FTDA and COALLIA's respective internal monitoring systems. For newly referred refugees, randomisation is based on the monthly lists provided by OFII to the AGIR programme, which are shared in advance with all operators and the pre-registered profiles of the refugees in the AGIR monitoring system. The randomisation is done after project managers have anonymised the data. Neither the researchers in charge of the evaluation nor the social workers at FTDA and COALLIA know the group assignment of any refugee.

The randomisation is stratified by *Groups of nationality* $\times$ *gender* to ensure balance between the four groups, as we anticipate differentiated effects of the intervention based on these characteristics. More precisely, we construct these strata by imposing at least the inclusion of four individuals in each (one per group). When *country* $\times$ *gender* stratification is not possible, we group remaining nationalities into broader groups based on region of origin (*e.g.*, Europe and Central Asia, Latin America and Caribbean, etc.) following the World Bank regional classification.

This stratification improves the precision of our treatment effect estimates. Specifically, if gender and region of origin are correlated with our outcomes of interest, the stratification helps reduce residual variance and increases the statistical power of the analysis. In turn, this enables a more accurate assessment of treatment effects within each subgroup.

The materials handed to participants, and the figures documenting who was seen and who consented, are collected in Appendix A.2.

Crossovers and spillovers. Both may occur, particularly with respect to the helpline treatment, as refugees may share the helpline number with relatives or acquaintances outside the programme. A key feature of our design is that we can identify and track such cases: when a refugee calls the helpline, we can determine whether they are enrolled in RISE and to which treatment group they belong.[26] We nonetheless adopt an intention-to-treat framework, analysing individuals based on their initial random assignment irrespective of actual exposure to the intervention.

[26] As of the July 2026 extraction, no crossovers have been recorded: all 22 recorded calls come from the two arms granted access to the number. We do however note two types of pure spillovers. First, individuals neither enrolled in RISE nor in AGIR have occasionally called the helpline to request information. Second, some partner health structures have reported visits from individuals who identified themselves as RISE participants but were not enrolled in the project. These pure spillovers do not affect our estimates but will be quantified at the end of the project to grasp the magnitude of this potential mechanism in the case of a scale up.

3.3 Data sources

We draw on a variety of complementary data sources. First, we use administrative records from the AGIR programme monitoring system. Second, we use original survey data collected from all participants who consented to the study, including a baseline survey (the endline survey will take place in early 2027). Third, we exploit data from partner mental health providers and the Paris public hospital network (*Assistance Publique – Hôpitaux de Paris*, AP-HP) to measure service utilisation, which constitutes our primary set of outcomes. Fourth, we *plan* to use additional administrative data from French public authorities to track individuals over the longer run on a various set of health, and labour market outcomes.
Finally, we use data collected through the helpline by peer-helpers over the entire duration of the project.

Partners' Monitoring Data. Data from FTDA and COALLIA's monitoring systems (ANITO and SILAO, respectively), to which we were granted secure private access, are used to implement the stratified randomisation, assess balance on baseline characteristics across treatment groups and to include control variables in robustness checks. This database provides detailed information on beneficiaries' demographic characteristics (nationality, country of birth, gender, age, marital status, household position, number of children, education, and language skills), migration background (years since arrival in France, time since signing the CIR, and time since enrolment in the AGIR programme) and anonymised identifiers for the two caseworkers assigned to each beneficiary. It also allows us to track participants' exit from the programme, as well as the reasons for exit, including expiration of eligibility, transitions to stable employment and housing, exclusion, or transfer to partner programmes.

Access to these unique administrative records substantially reduces the need to collect information directly from refugees. However, since the data are recorded by caseworkers during regular interactions with beneficiaries, some pieces of information may be missing or incomplete. To mitigate this, we cross-check information across sources and implement additional data processing procedures. In particular, entry date to France, housing conditions and employment status at entry were recovered from the unstructured caseworker notes in order to recover missing values where possible.

Baseline Survey. The baseline survey consists of a short self-reported health questionnaire administered to refugees immediately after they consented to the study. The English version is provided in B.3. The questionnaire is self-administered on a tablet in the respondent's preferred language, prior to disclosure of their treatment group assignment.[27] It collects information on general health status and basic knowledge of the French healthcare system In addition, mental health is measured using items from the *International Trauma Questionnaire* (Cloitre et al., 2018), a validated instrument designed to assess symptoms associated with Post-Traumatic Stress Disorder (PTSD). This tool has been widely validated across populations and languages, including among refugees (Seiler et al., 2023). These measures do not provide clinical diagnoses but instead capture self-reported symptomatology consistent with the *International Classifica-*

[27] Respondents have the option of listening to the questions rather than reading them.

tion of Diseases (ICD-11) framework.

Endline Survey. The endline survey will consist of a short questionnaire collecting data on self-reported health status, healthcare utilisation, and socio-economic integration. The English version is available upon request to the authors. The questionnaire will be administered to all participants by phone, in each refugee's main language, in January 2027. It will allow us to estimate medium-run treatment effects and to expand the set of potential outcomes with additional detailed measures on mental healthcare access and socioeconomic integration.[28]

Mental Health Provider. Data on healthcare utilisation first come from the four main facilities serving refugees in Paris, namely COMEDE, Le Chêne et l'Hibiscus, Avicenne Hospital, and Centre Primo Levi. These facilities correspond to four of the six providers listed in the leaflet distributed to all participants and identified with a "RISE Partner" stamp. These facilities identify AGIR-enrolled refugees in their information systems and regularly transmit to the project managers individual-level data on contacts, including the number of appointments scheduled and attended. No information on the content of consultations is shared, ensuring the confidentiality of clinical details. Data sharing was governed by formal agreements with each facility for the duration of the project.[29]

AP-HP Data. The AP-HP hospital monitoring data (ORBIS) records any hospital-admission dates per individual, corresponding to every distinct admission to a hospital. Each record includes the patient's name, date of visit, and the name of the facility visited. Access to the AP-HP data was made possible through the partnership with Avicenne Hospital and the involvement of a member of Avicenne's medical staff among the project team.[30]Unfortunately, the information we had access to at this stage is limited to hospitals in the AP-HP (*Assistance Publique-Hôpitaux de Paris*), therefore not including health information outside the Paris area. Another limitation comes from the fact that we had access to the nature of the visit only for our partner Avicenne hospital. This implies that, for other structures of the Paris area, we cannot know whether the refugee came only once for an emergency, or multiple times for consultations, nor the reason of the consultations.[31]

Other Administrative Data. We plan to match participants with several sets of administrative data from different ministries, covering socio-economic, labour market, and health trajectories of participants using the refugee identifier. More specifically, we plan to study trajectories on residence-permit acquisition, and civic-integration, labour-market histories covering employment, unemployment and vocational training, and healthcare consumption together with the take-up of social and housing benefits. These sources would extend the follow-up beyond the horizon of the current results and allow medium-run treatment effects to be estimated on out-

[28]Socioeconomic integration variables, including employment, housing, social networks, and language acquisition, will draw on established survey instruments used in refugee studies, notably the IAB-BAMF-SOEP Survey of Refugees in Germany (Brücker et al., 2018).

[29]We cross-check our participant list against their records to make sure no visit was missed.

[30]The ORBIS dataset includes only the first time a person visited a Parisian hospital, if a person visited the same hospital more than once, only the first visit was recorded.

[31]It is worth noting that Avicenne is the AP-HP hospital most specialised in refugee mental health, to which refugees are typically referred when presenting mental health issues.

comes that our partner monitoring data cannot observe. Importantly, access to administrative healthcare consumption would allow precise measures of engagement with health and mental healthcare in a setting where individuals might choose to turn to non-partner structures. Access is not yet granted, so the results presented do not rely on them.

Helpline. Information collected by peer-helpers through the helpline will allow us to construct four measures of helpline take-up. By design, these outcomes are mainly observed for participants assigned to the helpline or peer-helper+helpline treatment arms. However, calls from participants in other groups or from non-participants may occur and would reflect potential spillovers.

3.4 Variables of interest

In this section we briefly describe the construction of our outcomes as well as the mental health measure derived from our baseline survey (for more details, see B.4). We characterise take-up variables into four broad categories: helpline outcomes, mental health partner structure outcomes, hospital outcomes, and any signal. Each outcome below is summarised here and defined in full, source by source, in Appendix B.1. We combine self-reported survey data[32] with administrative records from partner mental health facilities and other administrative sources. This multi-source approach allows us to capture both subjective measures of health and well-being, alongside objective indicators of healthcare utilisation, such as hospital visits and helpline calls.

Helpline Outcomes. We report the following variables: First, we define an indicator for whether a participant called the helpline. Second, we construct an indicator for whether the call was related to a mental health concern. Third, we define an indicator for whether the interaction resulted in a scheduled appointment arranged by the helpline staff. Finally, we construct an indicator for whether the participant attended the appointment scheduled through the helpline

Partner-Facility Outcomes. Our partner mental health structures regularly report whether any refugee enrolled in RISE visited their facility. We construct three variables from these data. First, an indicator for whether a participant visited any partner structure after their treatment date. Second, the total number of such visits. Third, the number of visits involving a consultation with a psychologist or psychiatrist specifically. These outcomes, together with the hospital and any-signal outcomes below, are estimated in Appendix D.1.

Hospital Outcomes. From the AP-HP administrative data (ORBIS), we construct two after-treatment variables: an indicator for any hospital visit after the treatment date, and the total number of such visits. We also construct the number of pre-treatment hospital visits, the number of hospital visits after treatment from which we derive information on whether the refugee had a hospital visit.

[32] Only at baseline at this stage of the paper, the endline survey will be implemented in January 2027

Any Signal. This binary variable equals 1 if at least one take-up signal is recorded for the refugee after their treatment date, that is, any call to the helpline, visit to a partner mental health structure, or hospital visit, and 0 otherwise.

Exit from the AGIR Programme. We record exit dates occurring before the administrative cap of 24 months, along with the reason for exit. These variables are of particular interest as a successful exit, defined as finding stable housing and employment, is precisely one of the outcomes our intervention aims to improve. The fourteen recorded motives and the way they are grouped into exit categories are detailed in Appendix B.1, and the estimates are reported in Appendix D.2.

Mental health variables. We use our baseline survey to construct multiple mental health variables, whose item-by-item distribution and diagnostic construction are documented in Appendix B.4, Table B4. As reported in Appendix Table B5, the PTSD module captures four symptom clusters: re-experiencing, avoidance, sense of current threat, and functional impairment. The first three clusters are each measured using two items assessing the extent to which respondents have been bothered by specific symptoms over the past month. Responses are recorded on a five-point Likert scale ranging from "Not at all" to "Extremely". A module is considered evaluable if at least one question within it has been answered.

In addition to these symptom measures, functional impairment is assessed through three additional items capturing the extent to which these difficulties affect respondents' daily lives, including their social interactions, work, or family responsibilities. A symptom is considered present if it is reported at least at a moderate level (≥ 2). Then, an individual is classified as meeting the criteria for PTSD when at least one symptom is present in each of the three core clusters and at least one indicator of functional impairment is met.[33]

Two appendix tables document how far item non-response constrains the diagnosis. Table B5 reports, for each diagnostic component, the share of respondents for whom that component can be evaluated at all, that is, for whom at least one of its items is answered. Evaluability is high and remarkably even across components, between 89.9% for negative self-concept and 93.2% for re-experiencing, so no single cluster drives the loss. The diagnosis nonetheless requires every core cluster to be evaluable simultaneously, which is why it can be established for 954 of the 1,135 refugees. Table B6 then reports the share of individuals diagnosed with PTSD by number of missing responses. Among individuals who answered all PTSD-related questions, 24.4% are diagnosed with PTSD. This share decreases systematically as the number of missing responses increases, already declining noticeably when only one question is left unanswered. A missing response indicates that the respondent chose not to answer a given question, which may itself signal discomfort with its content. As a consequence, PTSD cannot be assessed for a subset of

[33]Our questionnaire also includes additional items related to disturbances in self-organisation (DSO), which are typically used to construct a complementary measure of Complex Post-Traumatic Stress Disorder (CPTSD), capturing affective dysregulation, negative self-concept, and disturbances in relationships. However, as with PTSD, the identification of CPTSD requires combining these symptom domains with corresponding measures of functional impairment. In this study, we do not construct a full CPTSD measure. This choice reflects a trade-off between measurement completeness and respondent burden: extending the questionnaire to include all required items on sensitive topics could have reduced participation among refugees enrolled in the AGIR programme.

participants, likely leading us to underestimate its true prevalence in the sample.

The baseline questionnaire also includes a shorter set of items capturing key dimensions of emotional distress. In particular, we administer two items from the Affective cluster of the International Depression Questionnaire (Kroenke et al., 2003). Respondents were asked: *"Over the past two weeks, how often have you been bothered by the following problems?"*, using a six-point scale ranging from "Never" to "Almost every day". The two items are: (i) *"I find no interest or pleasure in doing things"*, and (ii) *" I have a feeling of sadness, depression or hopelessness"*. We take advantage of these two additional items to construct a depression score by summing these two items, yielding a measure ranging from 0 to 6, defined for respondents with non-missing answers to both questions. Higher values indicate more severe depressive symptoms.

Self-reported data on health status, healthcare utilisation, and socio-economic integration will be collected through an endline phone survey conducted in the refugee's main language in January 2027, and administered to all participants. This will allow us to estimate medium-run treatment effects and to expand the set of potential outcomes with additional detailed measures on mental healthcare access and socioeconomic integration.

4 Sample description

The final sample in Paris comprises 1,135 participants who consented to take part in the project, with the distribution across treatment arms and the control group reported in Figure 3.2. The sample is composed of relatively young adults, with an average age at entry of 31 years, and a minority of women (17.5%). Tables C2 and C3 report descriptive statistics for the final sample. Participants are predominantly refugees (92.7%), with a small fraction being relatives enrolled in the AGIR programme by virtue of their connection to a beneficiary.[34] They have arrived in France relatively recently, with an average of 3.2 years since arrival. At the time of treatment, individuals have been enrolled in the AGIR programme for an average of 11.8 months and 15.4 months since signing the CIR.[35]

As shown by Figure 3.3, the sample is dominated by a few large nationality groups, with Afghan refugees representing the largest share (35.2%), followed by Bangladeshi (22.5%) and Sub-Saharan African (excluding Côte d'Ivoire) participants (15.8%). Smaller groups include Sri Lankan (7.8%), Ivorian (5.2%) and Middle Eastern and North African participants (4.7%). The remaining categories, including Latin America and Europe and Central Asia, represent relatively small shares.

4.1 Refugees' Health Profile

Mental health. Figure 3.4 reports the average share of respondents by demographic and socioeconomic subgroups who meet a positive PTSD diagnosis as defined in Section 3.4. On average, 21.4% of the 954 refugees for whom the diagnosis can be established meet the criteria for

[34] Parents or partners of a refugee.
[35] All characteristics are constructed using non-missing observations for each variable.

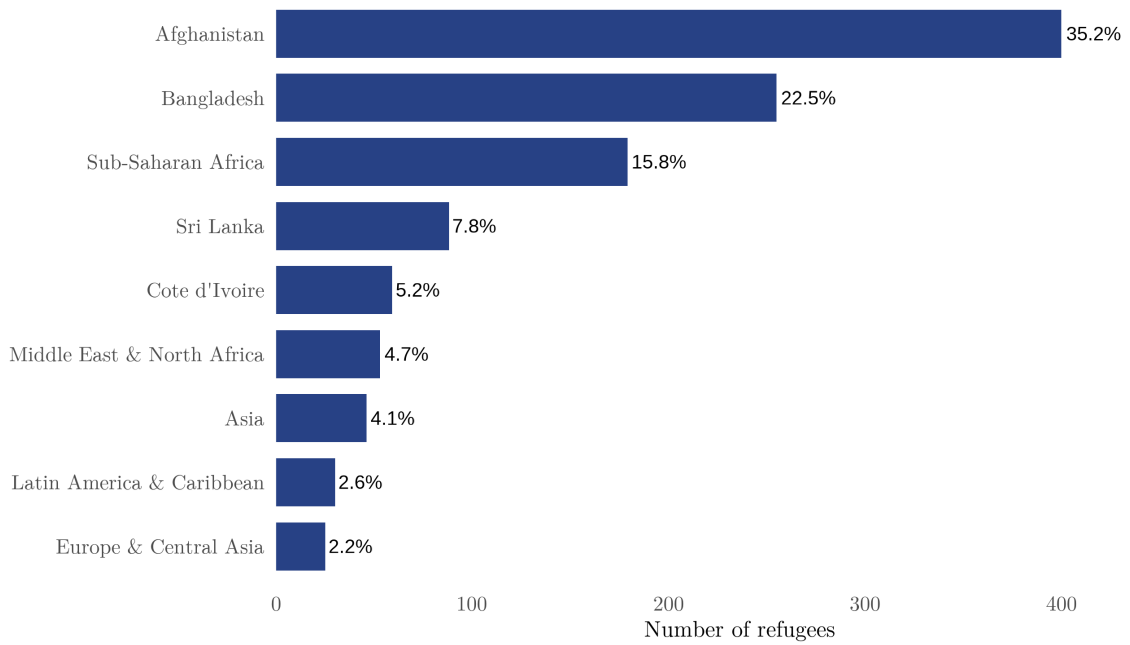

Figure 3.3: Breakdown of nationalities

Notes: This figure depicts the distribution of individual nationalities and nationality groups in our sample, which form the basis of the randomisation strata.

PTSD.[36] While Foged et al. (2026) find a PTSD prevalence of 30% in their study on newly arrived Ukrainian refugees, they note that prevalence varies considerably across refugee populations depending on context and degree of trauma exposure. The rate observed in our sample is substantial and consistent with existing evidence documenting high levels of trauma exposure among refugee populations.

Mental health heterogeneity. We observe some heterogeneity across subgroups within our sample. In particular, PTSD prevalence is higher among female participants (25.9%) than among males (20.5%). Differences also emerge across employment status at entry, with individuals in precarious employment having the highest prevalence (22.9%), against 22.5% among those in stable employment and 20.7% among the unemployed. Housing conditions at entry also correlate with PTSD prevalence. Participants in more precarious living situations, such as reception centres or hotels (25.3%), social or temporary housing (24.5%), and homelessness (25.0%), display higher rates than those in more stable arrangements, such as private housing (15.4%). This comparison should however be interpreted with caution, as only 15 individuals reported private or stable housing at entry.[37] This provides suggestive evidence of a positive association between mental health and precariousness.

General health and knowledge. Our baseline survey also includes questions on general

[36] Figure B1 in the Appendix reports the distribution of the ITQ PTSD symptom score, constructed by summing responses across all items to obtain a continuous measure of PTSD severity.

[37] While the PHQ-2 depression screen is strongly correlated with PTSD, its baseline prevalence equals 26.5% and exhibits less heterogeneity across subgroups, as shown in Appendix-Figure B2. Notably, among individuals who meet the PTSD criteria, 58.0% also screen positive for depression, against 18.6% among those who do not.

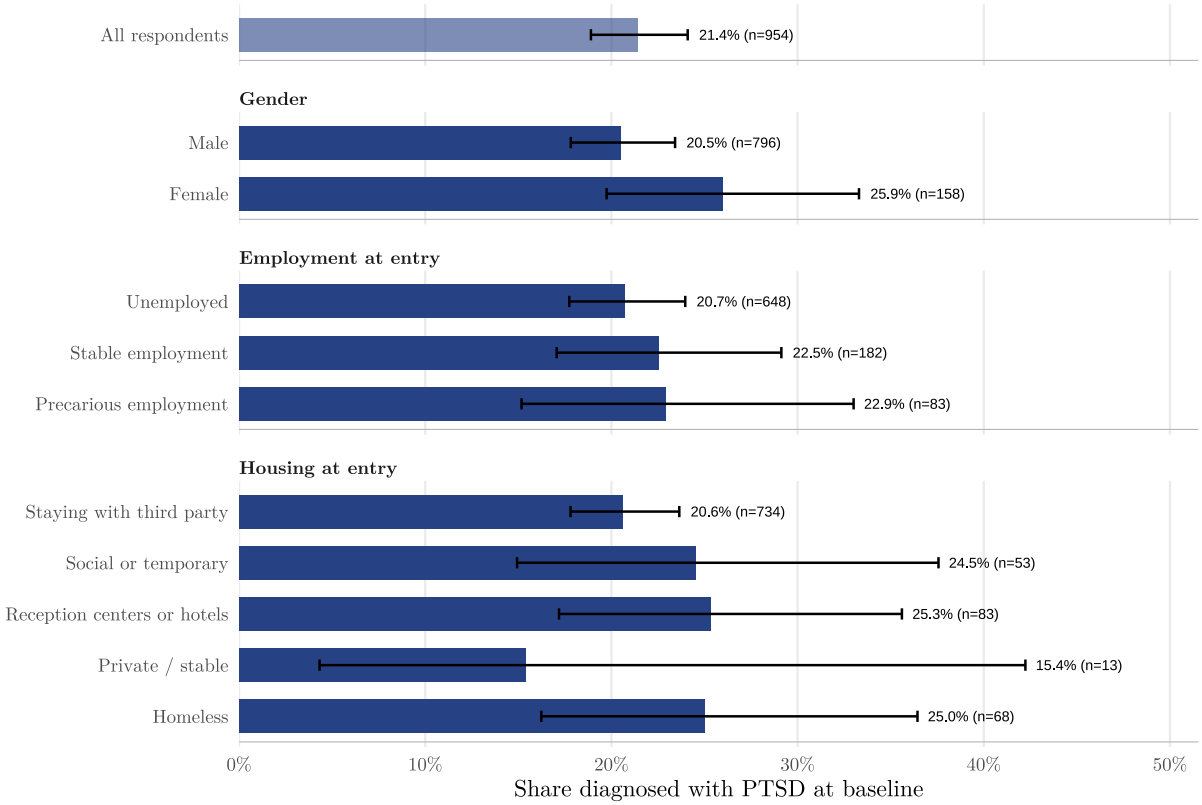

Figure 3.4: Baseline PTSD prevalence, by subgroup

Notes: Each bar reports the share of baseline respondents in the indicated subgroup who meet the ICD-11 PTSD diagnostic criteria (endorsement scored ≥ 2 on at least one item in each of re-experiencing, avoidance and sense-of-threat clusters, and on at least one of the three functional-impairment items P_7-P_9). Horizontal whiskers are Wilson 95% confidence intervals around the subgroup share and are therefore not centred on the observed proportion (they shrink slightly toward 0.5, especially in small cells). Subgroup N in parentheses is the number of diagnosable respondents in that cell.

health. We find that 62.5% of respondents report being in very good or good health, 30.9% in average health, and 6.6% in poor or very poor health.[38] Regarding satisfaction with current health, 72.3% of respondents report being quite or very satisfied, while 27.7% report being not or not really satisfied. Finally, only 24.3% of respondents report knowing the French healthcare system well or very well, while 75.7% report limited or no knowledge. Consistent with this, access to care appears challenging. Indeed, 55.1% of respondents report that obtaining an appointment is rather (38.3%) or very difficult (16.8%), compared to 44.9% who report that it is easy (31.3%) or very easy (13.6%). This supports the relevance of interventions aimed at reducing organisational barriers related to the complexity of navigating a healthcare system that is unfamiliar to refugees.

[38] Among individuals admitted on humanitarian grounds, the ELIPA2 survey reports that four years after receiving their first residence permit, 77% are in very good or good health, 15% in fair health, and 8% in poor or very poor health (ELIPA2, 2024a).

4.2 Refugees' Economic and Housing Integration

Ultimately, the intervention aims to improve refugees' economic integration by facilitating engagement with mental healthcare pathways, which may, in turn, support progress toward the AGIR programme's core objectives of securing stable housing and employment. This motivation is particularly relevant given the high level of socioeconomic vulnerability in our sample.

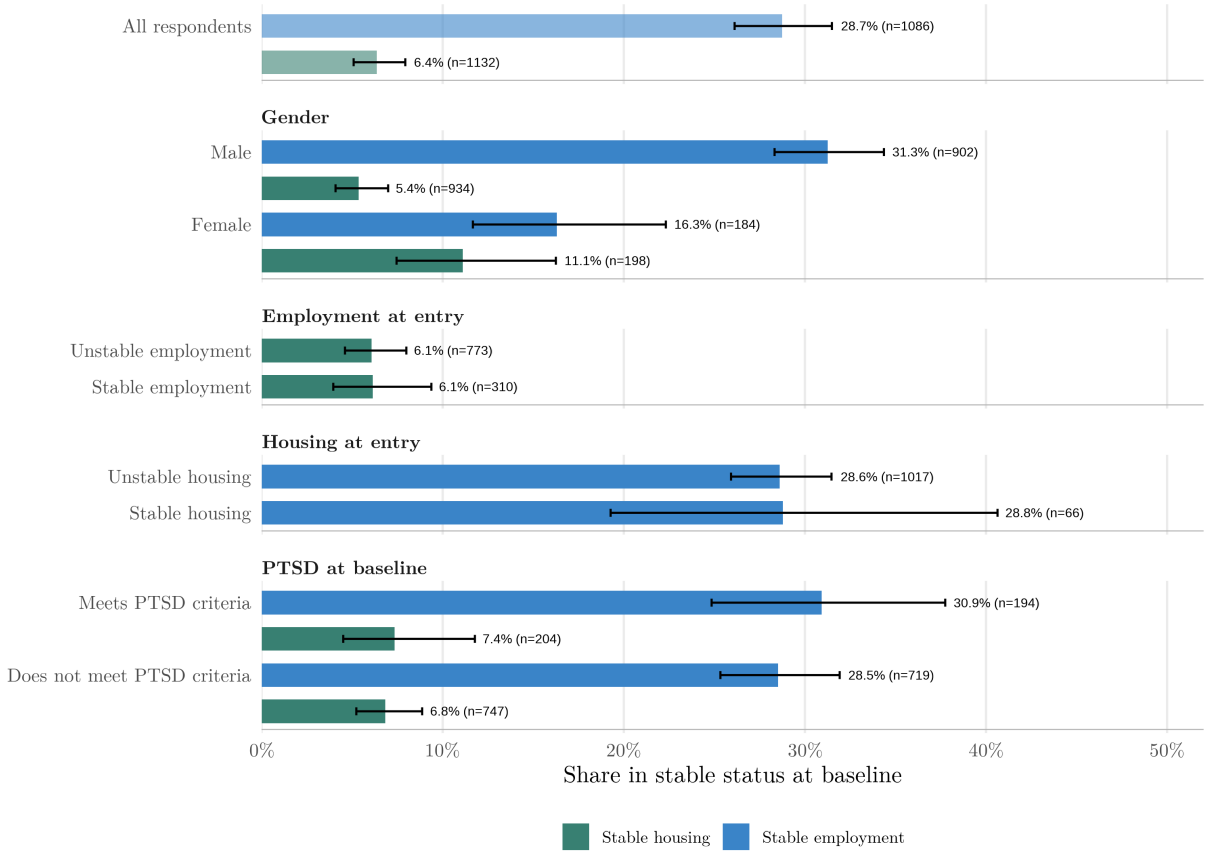

Figure 3.5: Baseline prevalence of unemployment, by subgroup

Notes: Within each row, paired bars report the share of respondents in stable housing (green) and the share in stable employment (blue), either in the full sample (top row) or in a given subgroup (remaining rows). Stable housing groups private or stable housing and social or temporary housing; all other modalities are classified as unstable. Stable employment pools refugees reporting paid employment and those in vocational training at entry; unstable employment is the complement (unemployed). Horizontal whiskers are Wilson 95% confidence intervals around the subgroup share and are therefore not centred on the observed proportion (they shrink slightly toward 0.5, especially for small cells). Subgroup N, in parentheses next to each bar, is the number of respondents with a non-missing value for the corresponding outcome. Within the Employment (resp. Housing) facet, the stable-employment (resp. stable-housing) bar is mechanically 100% in the stable level and 0% in the unstable level; those bars are retained to preserve visual symmetry across dimensions.

Figure 3.5 shows that only 28.7% of participants are in employment or vocational training at entry, while the remaining 71.3% are unemployed or outside the labour market. These disparities are especially pronounced across gender, with 16.3% of women in employment or training against 31.3% of men. Housing conditions are similarly precarious. Only 6.4% of participants report stable housing at entry (private or social/temporary housing), implying that the vast majority

are in unstable living arrangements, including homelessness, accommodation with third parties, or residence in reception centres or hotels. Interestingly, women appear less exposed to housing instability than men, which may reflect gender-specific allocation mechanisms in access to public housing or support services. Finally, the figure does not reveal a strong association between employment and housing conditions at baseline, as differences across subgroups remain limited along these two dimensions.

5 Empirical Strategy

This section presents the empirical strategy used to estimate the effects of the three treatment arms on healthcare utilisation and economic integration outcomes. We first assess the validity of the randomisation by examining to which extent the distribution of covariates is similar across levels of treatments. We then present our main estimation framework and benchmark specification.

5.1 Balance between Groups

The validity of our research design relies on the assumption that random assignment generates comparable groups across treatment arms. To assess whether randomisation achieved balance, we estimate the following regression:

$$C_i = \alpha + \beta_1 G_{1,i} + \beta_2 G_{2,i} + \beta_3 G_{3,i} + \sum_i \delta_s \mathbf{1}[i \in s] + \epsilon_i, \tag{3.1}$$

where C_i denotes a baseline characteristic for individual i, and $G_{x,i}$, with $x \in \{1, 2, 3\}$, dummy variables equal to one if refugee i was assigned to group x, and zero otherwise. $\mathbf{1}[i \in s]$ are dummies for the randomisation strata (at the *gender* $\times$ *group of nationality* level). Figure 3.6 reports standardised differences in baseline characteristics between each treatment arm and the control group. Differences are small in magnitude, and generally not statistically significant indicating a high degree of balance across groups.

To assess whether existing differences reflect any systematic imbalance, we report in C pairwise balance tests between groups. Then, we further adjust for multiple hypothesis testing using the stepwise procedure proposed by Romano and Wolf (2005). Reassuringly, after this adjustment, we fail to reject the null hypothesis of equality across groups for all baseline characteristics.

5.2 Intention-to-Treat estimates

Our benchmark specification estimates the causal effect of random assignment to each treatment arm on outcome Y_i by ordinary least squares:

$$Y_i = \alpha + \beta_1 G_{1,i} + \beta_2 G_{2,i} + \beta_3 G_{3,i} + \sum_i \delta_s \mathbf{1}[i \in s] + \epsilon_i, \tag{3.2}$$

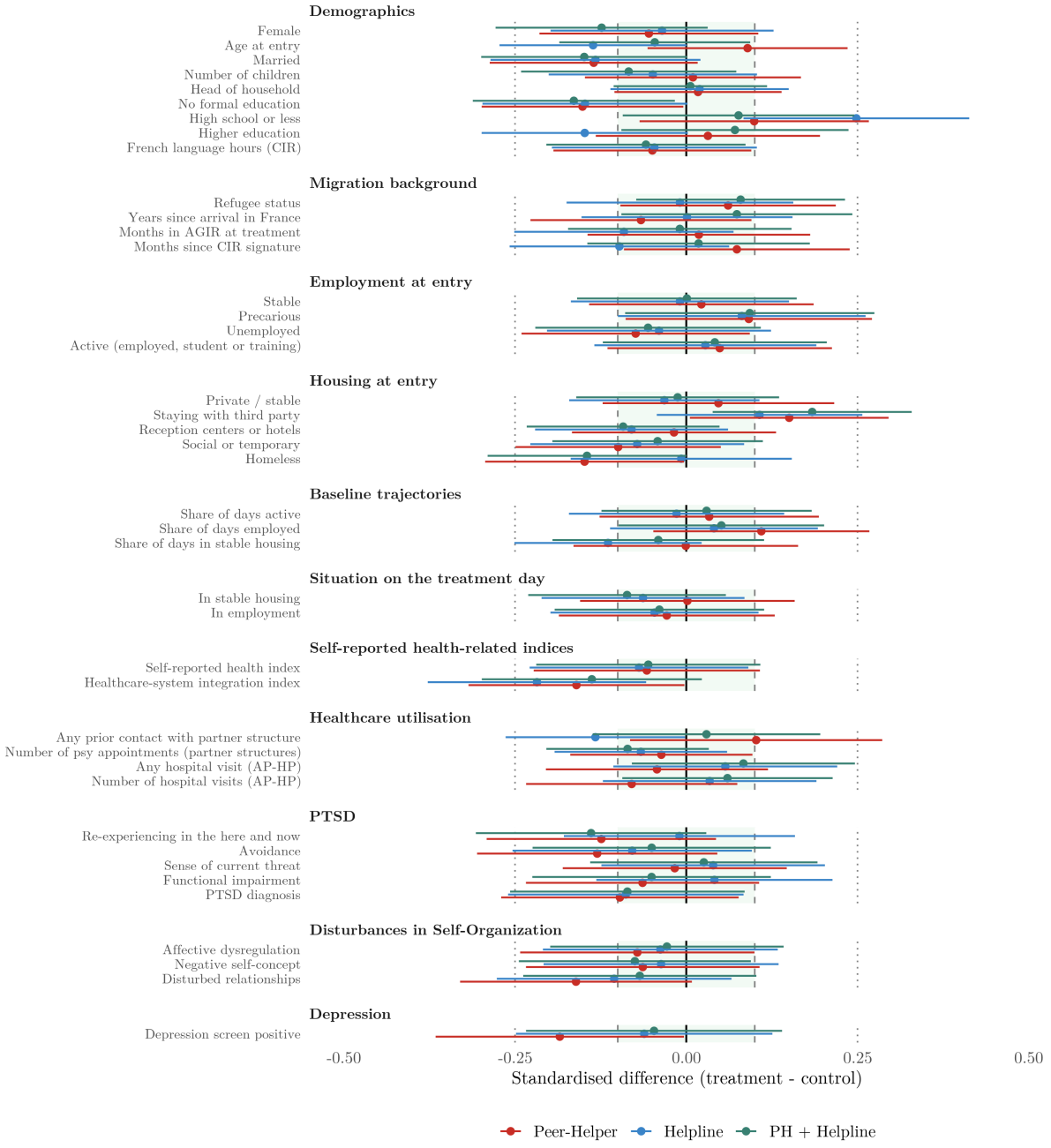

Figure 3.6: Standardised differences across treatment arms

Notes: For each baseline characteristic and each treatment arm, the figure plots the standardised difference from the Control group, computed as the arm-minus-Control mean divided by the Control-group standard deviation. Horizontal bars are 95% confidence intervals. Standardised differences between -0.1 and 0.1 (shaded band) are typically considered negligible in randomised-controlled-trial balance diagnostics. Variables are grouped by theme (demographics, migration history, socioeconomic status at entry, baseline mental health) and sorted within each group. The sample is the full four-arm sample; cells with missing values are dropped per variable.

where $G_{k,i}$ is a dummy variable equal to one if refugee i was assigned to treatment arm $k \in \{1, 2, 3\}$ and zero otherwise, where $k = 1$ denotes the Peer-Helper group, $k = 2$ the Helpline group, and $k = 3$ the combined treatment. $\mathbf{1}[i \in s]$ are dummies for the randomisation strata. The coefficients $\beta_1, \beta_2, \beta_3$ give the causal impact of each arm relative to the control group. ϵ_i is

the error term, with robust standard errors at the individual level.

5.3 2SLS estimates

The previous specification investigates the effects of the intervention on take-up, that is, the probability of engaging in mental health care. We extend this analysis by estimating the average treatment effect on compliers (LATE) on employment and housing integration, using random assignment as an instrument for take-up. It allows us to assess whether potential improvements in mental health translate into better socioeconomic outcomes, such as employment and housing stability. To do this, we estimate the following linear model:

$$I_i = \xi + \mu \widehat{\mathrm{Y}_i} + \sum_i \delta_s \mathbf{1}[i \in s] + \nu_i, \tag{3.3}$$

where I_i is an outcome of economic or housing integration and $\widehat{\mathrm{Y}_i}$ is obtained from the first stage equation estimated in Equation 3.2. The coefficient of interest is μ, which identifies the causal impact of engaging with the intervention on compliers, *i.e.*, refugees who take up the intervention because they are assigned to the treatment arm. It requires monotonicity, which is plausible here since no arm can discourage a refugee from seeking care, and it requires that assignment affects integration only through the care contact we instrument. This is potentially a fragile assumption as contact with the health system may trigger administrative steps, referrals or entitlements that can also affect employment and housing, as well as mental health because the peer-helper conversation may for example convey information or encouragement of its own. Both violations push towards overstating the size of the estimates. We thus report them below as a discussion piece and as an upper bound, not as our preferred evidence.

5.4 Heterogeneous Effects

We extend our analysis by exploring heterogeneous treatment effects across baseline characteristics. To do so, we estimate the following specification:

$$Y_i = \alpha + \boldsymbol{\beta'} \boldsymbol{G_i} + \eta Z_i + \boldsymbol{\theta'}(\boldsymbol{G_i} \times Z_i) + \boldsymbol{X}_i + \sum_i \delta_s \mathbf{1}[i \in s] + \nu_i, \tag{3.4}$$

where Y_i denotes the outcome of interest, $\boldsymbol{G_i}$ is a vector of treatment arm indicators for individual i, $\boldsymbol{X}_i$ is the vector of baseline covariates, and Z_i is the variable used to assess heterogeneity in treatment effects. $\boldsymbol{\theta'}$ is the vector of coefficients of interest, which captures how the effect of being in each treatment group G_i, varies depending on the value of the variable Z_i. $\mathbf{1}[i \in s]$ are dummies for the randomisation strata.

While treatment effects may plausibly vary along several dimensions, the stratified randomisation by gender and origin ensures sufficient representation within each stratum to credibly examine heterogeneity along these margins. At this stage, we only focus on whether the effects of the interventions differ by gender, but plan to extend this analysis by groups of origin.

Indeed, cultural attitudes toward mental health, and the associated stigma, vary between societies, which can lead to differences in treatment effects. In contexts where stigma is particu-

larly strong, the peer-helper intervention may have a significant impact, as it directly challenges misconceptions about mental health. Gender may also shape responses to the intervention as men and women can face different barriers to accessing care, show different prevalence rates of mental health conditions, and experience varying levels of stigma related to seeking support. In particular, female refugees may have been exposed to specific forms of trauma, such as sexual violence, and may encounter distinct cultural constraints when discussing mental health. As a result, their interactions with peer-helpers could either increase or limit their comfort in addressing sensitive issues, potentially leading to stronger, or, in some cases, weaker treatment effects.

A second dimension of heterogeneity is internal to the intervention itself. The peer-helper arm bundles a message with a messenger, and its rationale is that stigma and mistrust decreases because the person delivering the message shares the participant's migration background and language. If that is the mechanism, then who delivers the conversation should matter, and effects should concentrate on particular peer helpers. Instead, if the content of the conversation is what drives the effect, it should be broadly similar across peer-helpers. Because assignment to a peer helper is not random, we examine this margin descriptively, using a counterfactual assignment methodology described in Appendix E.2. The interpretation is further constrained by what a peer helper is in our setting. Assignment follows the participant's language and gender, so each peer serves a distinct population, and the cell that identifies a given peer also identifies a nationality group, a language, and whatever stigma attaches to mental health in that community. A contrast across peers is therefore a contrast between peer-times-clientele combinations, and the messenger effect we are after cannot be separated from the composition of the group they serve, except among peers who work in the same language and whose assignment is effectively driven by availability alone.

6 Results

6.1 Results on take-up, engagement with mental healthcare and hospital visits

This section reports the effects of each treatment on the intermediary outcomes[39]. We first focus on a summary measure that captures whether the intervention generates any detectable effect, defined as an indicator equal to one if at least one positive signal is observed across the set of outcomes considered (Figure 3.7). We then turn to outcome-specific results, organised into three categories: helpline outcomes (Figure 3.8), partner-facility outcomes (Figure 3.9), and hospital-based outcomes (Figure 3.10). All figures report treatment effects estimated without additional controls, but including *Groups of nationality* × *gender* strata fixed effects.[40]

Benchmark results. Figure 3.7 reports the effects of the treatments on the summary indicator

[39] As outcomes are collected continuously over time, the estimates presented in this section should be interpreted as preliminary evidence reflecting a snapshot of treatment effects as of 8 July 2026 for the three healthcare data sources and 28 July 2026 for the AGIR monitoring data. These effects may evolve as data collection progresses.

[40] Corresponding tables, which additionally report tests of equality of treatment effects across arms as well as joint significance tests and Benjamini–Hochberg adjusted q-values are reported in D.5.

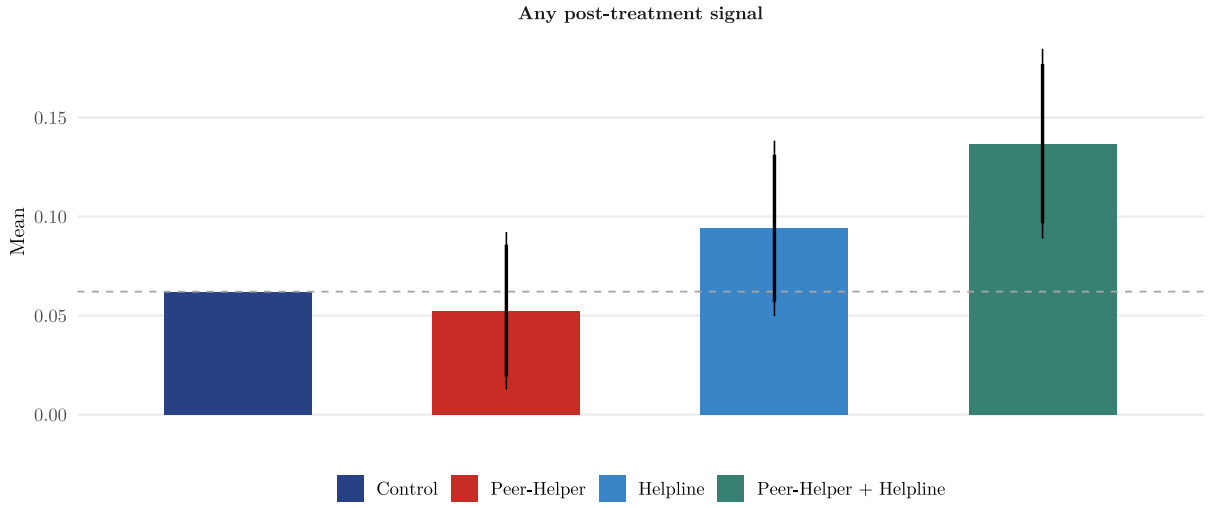

Figure 3.7: Effect on any post-treatment signal, by treatment arm

Notes: Bars report the post-treatment mean of the *any post-treatment signal* indicator in each treatment arm, where the indicator equals one if the refugee has at least one post-treatment event in any of the three data sources (partner-structure appointment, hospital admission, helpline call) and zero otherwise. The *control* bar is the unadjusted control-group sample mean. Each treated-arm bar equals the control-arm mean plus the intention-to-treat OLS estimate from a pooled three-arm regression with strata fixed effects (interaction of nationality group and gender) and HC1 heteroskedasticity-robust standard errors, together with its 90% (thick) and 95% (thin) confidence intervals. The control bar has no confidence interval as it serves as the regression's reference category. The dashed horizontal line marks the control mean. Details of the regression estimates are in Appendix Table D1.

capturing whether any post-treatment signal is observed and zero otherwise. The figure shows a clear gradient across treatment arms. The peer-helper-only group displays no significant difference relative to the control group. On the contrary, individuals assigned to the helpline treatment are more likely to exhibit at least one post-treatment signal. This effect seems to be further amplified for individuals assigned to the combined treatment, who exhibit the largest increase in the probability of any signal.

The estimated effects are modest in magnitude. The helpline-only treatment raises the probability of observing any signal by 3.2 percentage points, relative to a control mean of 6.2%, an effect that is not statistically significant. The combined treatment raises it by 7.5 percentage points, an effect significant at the 1% level which more than doubles the control-group rate.

The corresponding regression results are reported in Table D1 and provide additional insight into these patterns. They confirm that the effect of the combined treatment is statistically significant and robust to multiple hypothesis testing, as indicated by the Benjamini–Hochberg adjusted q-values. The helpline-only treatment also generates positive but not significant effects, although these are smaller and less precisely estimated. Pairwise comparisons further indicate that the helpline treatment significantly outperforms the peer-helper-only treatment, and that the combined treatment also outperforms the peer-helper-only treatment. However, differences between the combined treatment and the helpline-only treatment are smaller and not consistently statistically significant. Thus, it provides little support for complementarity between the peer-helper and helpline components at this stage.

Overall, this first set of results suggests that the helpline component is the main driver of the observed effects, while the peer-helper component alone does not appear to increase engagement. These findings point to the importance of addressing institutional barriers, such as difficulties in navigating the healthcare system and booking appointments, in improving access to mental healthcare among refugees, rather than relying only on peer-based discussions to reduce stigma and informational constraints. One alternative interpretation is that, although peer support is delivered in the participants' native language and by trained peer-helpers with experience in mental health support, in a supportive and culturally appropriate environment, such interactions may not be sufficient to overcome the substantial stigma surrounding mental health documented in the literature. The remainder of this section examines these patterns in more detail by group of outcomes.

Helpline. Figure 3.8 reports the effects of the treatment on outcomes measured through the helpline. These effects can be interpreted as measures of take-up, *i.e.*, the extent to which treated individuals respond to the intervention. We first observe that the intervention has the intended effect: individuals assigned to a treatment arm that includes access to the helpline are significantly more likely to call it. This effect is statistically significant for both the helpline-only group and the combined treatment group. While the point estimate is slightly larger for the combined treatment, the difference between the two groups is again not statistically significant. In terms of magnitude, the effects remain relatively modest, corresponding to increases of 3.0 percentage points for the helpline-only arm and 4.3 percentage points for the combined arm in the likelihood of calling the helpline.

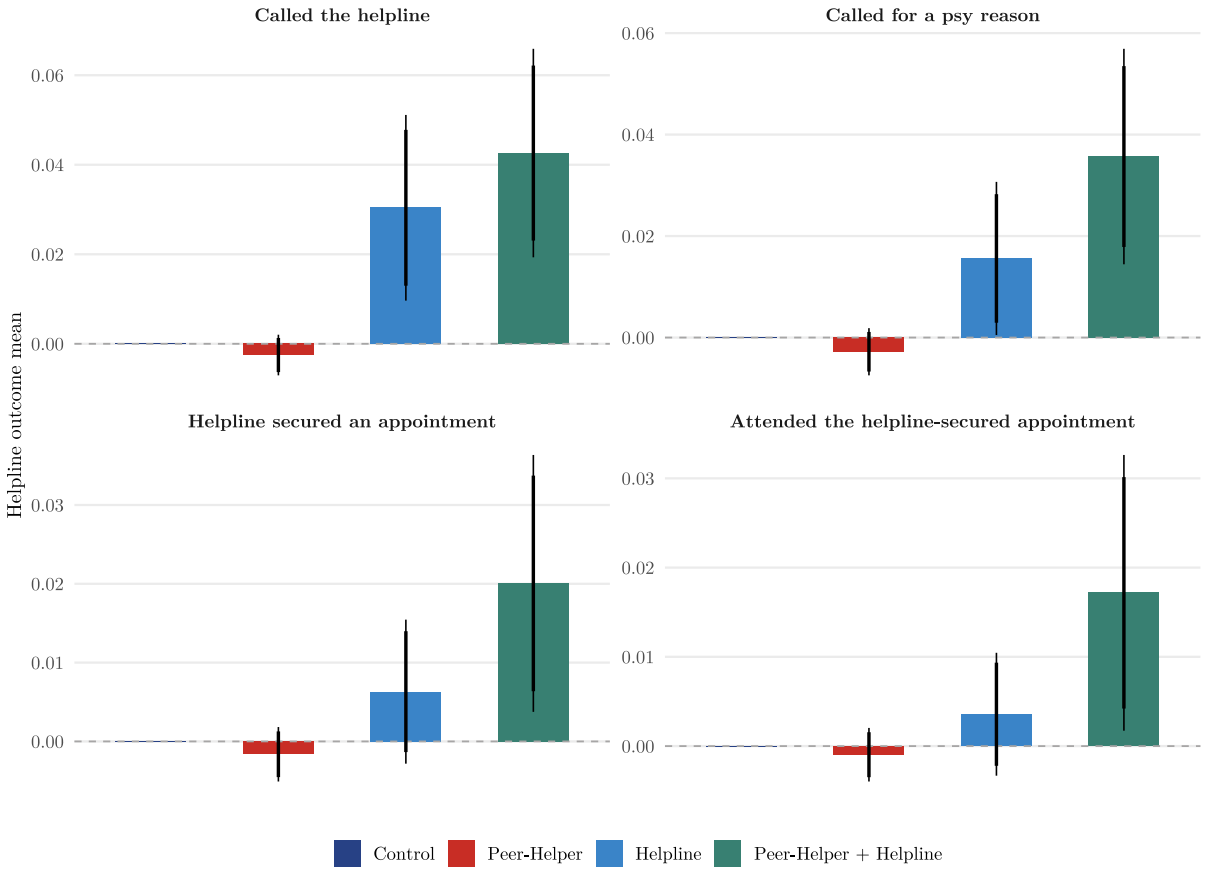

Figure 3.8: Effects on helpline outcomes, by treatment arm

Notes: Bars report the post-treatment mean of each of the four helpline-engagement indicators in each treatment arm. The outcomes, all binary, are *called the helpline* (at least one matched call to the dedicated helpline number), *called for a psychiatric reason* (call flagged by the mediator as motivated by a mental-health concern), *helpline secured an appointment* (mediator logged a downstream partner-structure appointment during or immediately after the call), and *attended the helpline-secured appointment*. Refugees in the *Control* and *Peer-Helper* arms are not given the helpline number by design, so their bars are mechanically at zero on all four outcomes; any non-zero value observed there would reflect a spillover of the number beyond the two assigned arms. The *Control* bar is the unadjusted Control-group sample mean. Each treated-arm bar equals the Control-arm mean plus the intention-to-treat OLS estimate from a pooled three-arm regression with strata fixed effects (interaction of nationality group and gender) and HC1 heteroskedasticity-robust standard errors, together with its 90% (thick) and 95% (thin) confidence intervals. The Control bar has no confidence interval as it serves as the regression's reference category. The dashed horizontal line marks the Control mean of each outcome. Details of the regression estimates are in Appendix Table D4.

The two arms differ, however, in what follows the call. In the combined arm, the intervention raises the probability that a mediator secures an appointment by 2.0 percentage points and that the appointment is attended by 1.7 percentage points, so that close to half of the induced calls translate into a scheduled appointment and almost all scheduled appointments are attended. In the helpline-only arm, the corresponding effects are smaller and not significant, at 0.6 and 0.4 percentage points respectively, so that only about a fifth of the induced calls reach the appointment stage. Conditional on a call being converted, therefore, entry into a mental healthcare pathway is effective. The difference between the two arms in conversion rates is not statistically significant, however, so we do not read it as evidence that the peer-helper component is required

at this stage. Finally, we find no significant effect of the peer-helper-only treatment on helpline use. This provides evidence against the presence of crossovers. In particular, it suggests that individuals not assigned to the helpline treatment did not access it indirectly through social interactions with treated participants.

Mental Health Structures. Figure 3.9 reports the effects of the treatment on outcomes measured through partner mental health facilities. These outcomes capture actual engagement with formal care, including whether individuals obtain an appointment, as well as the number of appointments scheduled. Overall, we find almost no statistically significant effects. The only exception is the probability of having at least one appointment in the combined arm, which increases by 1.3 percentage points against a control mean of 0.7%, significant at the 10% level once covariates are included. On the two count outcomes, point estimates are negative for all three arms and imprecisely measured, with wide confidence intervals that comfortably include zero, and the joint test that the three arm coefficients are zero is never rejected. In terms of magnitude, the estimated effects are modest. The largest point estimates correspond to increases of a few percentage points in the probability of having at least one appointment, and small increases in the number of appointments. These patterns are consistent with the relatively low take-up documented previously. Comparing treatment arms, the combined treatment outperforms the peer-helper-only treatment on the extensive margin only, and only marginally so once covariates are included. Similarly, the helpline-only treatment does not appear to generate meaningful changes in these outcomes. If anything, when appointments are observed, they do not appear to be systematically related to mental health care. This may reflect the broader health vulnerability of the study population and suggests that the intervention may have encouraged some individuals to engage with partner organisations that provide comprehensive support beyond mental health services.

Hospital visits. Figure 3.10 reports the effects of the treatment on hospital-based outcomes, using administrative data. In contrast to the previous results, we find evidence of positive effects for the combined treatment. Individuals assigned to both the peer-helper and helpline interventions are more likely to visit a hospital and exhibit an increase in the number of hospital visits. By contrast, the helpline-only and peer-helper-only treatments do not appear to generate effects of comparable magnitude. This suggests that the combination of informational support and facilitated access may be necessary to produce a meaningful impact on healthcare utilisation.

The effects are modest in absolute size, 4.5 percentage points for the probability of any visit and 0.05 additional visits, but large relative to a control mean of 5.5%, which they raise by about four fifths. They are also more precisely estimated and more consistent across outcomes and specifications than those observed for partner-facility utilisation, although they do not survive the Benjamini-Hochberg adjustment within the family, and the pairwise test rejects equality with the peer-helper-only arm at the 5% level. Interestingly, these hospital visits do not appear to be directly mediated by our helpline service. This suggests that the intervention may operate through a salience channel. While it does not necessarily translate into care within the RISE programme as reported in Figure 3.9, it may nonetheless encourage individuals to seek care independently in the broader healthcare system. This suggests implications for the scalability

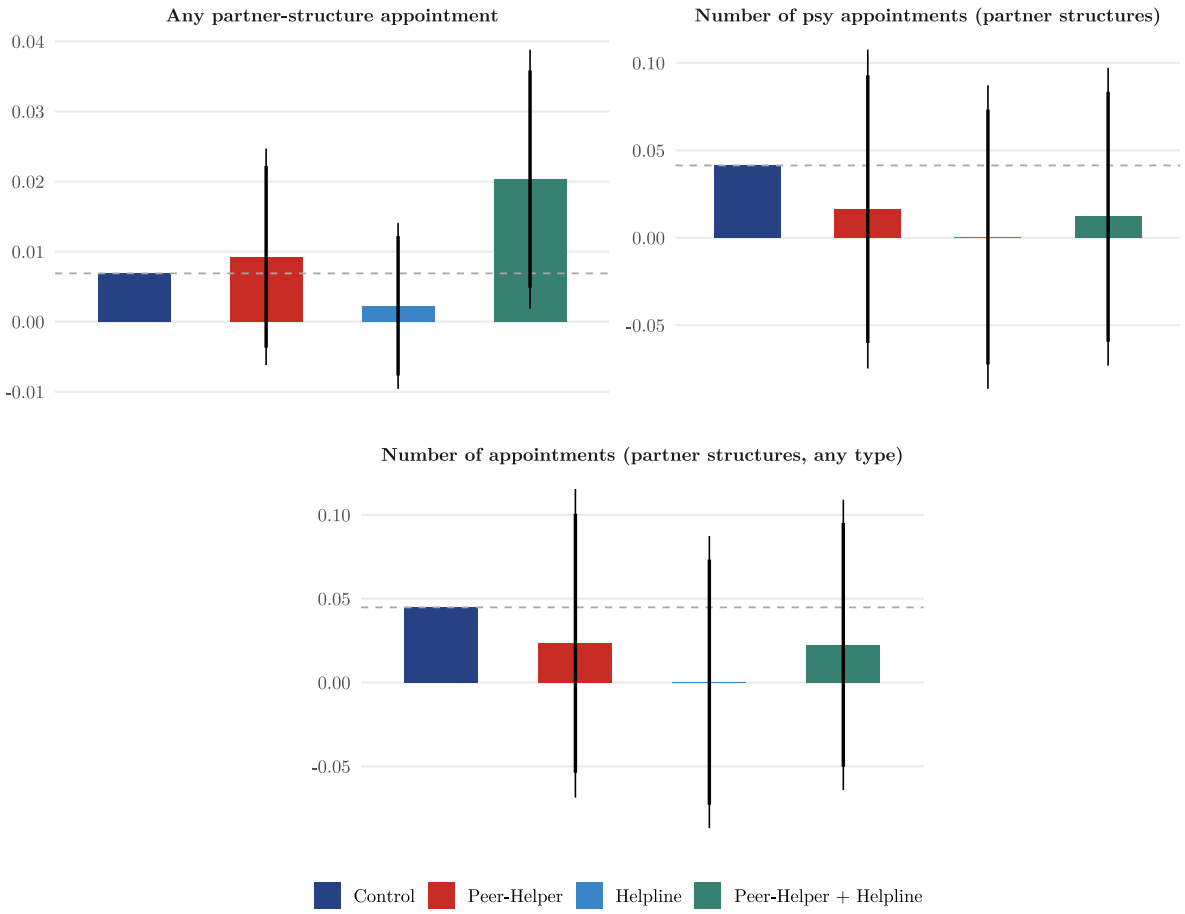

Figure 3.9: Effects on partner-structure outcomes, by treatment arm

Notes: Bars report the post-treatment mean of each of the three partner-structure outcomes in each treatment arm. The outcomes are *any partner-structure appointment*, an indicator equal to one if the refugee has at least one recorded appointment with any partner mental-health structure dated strictly after the treatment date and zero otherwise, and two counts aggregated from the partner-structure monitoring data: the number of psychiatric appointments and the number of appointments of any type (psychiatric, general-medical or administrative). The *Control* bar is the unadjusted Control-group sample mean. Each treated-arm bar equals the Control-arm mean plus the intention-to-treat OLS estimate from a pooled three-arm regression with strata fixed effects (interaction of nationality group and gender) and HC1 heteroskedasticity-robust standard errors, together with its 90% (thick) and 95% (thin) confidence intervals. The Control bar has no confidence interval as it serves as the regression's reference category. The dashed horizontal line marks the Control mean of each outcome. Details of the regression estimates are in Appendix Table D3.

of the intervention. In particular that its effects are not confined to a limited set of partner organisations and may be less dependent on the sustainability of project-specific funding, as refugees appear to engage with the broader public healthcare system.

Robustness to additional controls. For each outcome, the corresponding table in D.5 reports three nested specifications, reflecting increasingly rich sets of controls: (i) strata fixed effects only; (ii) strata fixed effects plus the full set of baseline covariates;[41] and (iii) strata fixed effects

[41] For every covariate with missing values we include a paired missingness indicator and recode the missing cells to a neutral placeholder, so that no observation is dropped from the regression while the coefficient on the original covariate is identified only on the observed sub-sample.

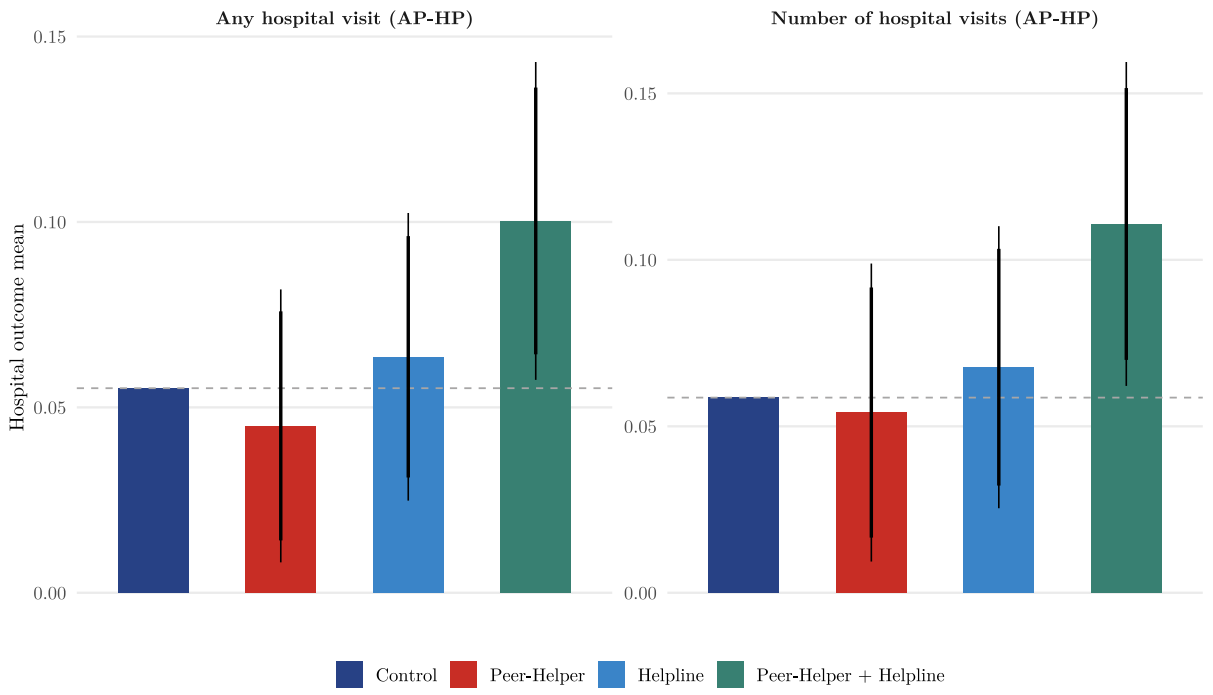

Figure 3.10: Effects on hospital outcomes, by treatment arm

Notes: Bars report the post-treatment mean of each of the two hospital-utilisation outcomes in each treatment arm. The outcomes are *any hospital visit*, an indicator equal to one if the refugee has at least one AP-HP ORBIS hospital admission dated strictly after the treatment date and zero otherwise, and *number of hospital visits*, an integer count of such admissions. The *Control* bar is the unadjusted Control-group sample mean. Each treated-arm bar equals the Control-arm mean plus the intention-to-treat OLS estimate from a pooled three-arm regression with strata fixed effects (interaction of nationality group and gender) and HC1 heteroskedasticity-robust standard errors, together with its 90% (thick) and 95% (thin) confidence intervals. The Control bar has no confidence interval as it serves as the regression's reference category. The dashed horizontal line marks the Control mean of each outcome. Details of the regression estimates are in Appendix-Table D2.

plus the subset of covariates selected using double-selection LASSO, following Belloni et al. (2014). Irrespective of whether controls are included and of the method used to construct them, our main conclusions remain unchanged, providing reassuring evidence on the validity of the randomisation. Furthermore, our estimates remain robust to the combined use of strata fixed effects and of fixed effects for the employment and housing social workers, reported in Tables D8 to D12.

Taken together, our results suggest that, at this stage of the intervention, treatment effects remain relatively modest. We find no detectable short-run impact of the peer-helper treatment on mental health service utilisation, which may reflect either limited effectiveness of this form of mediation or the persistence of strong stigma within this population that cannot be easily alleviated through a unique short interaction with a peer-helper. The helpline is used when it is offered, raising the probability of a call by 3.0 percentage points on its own, but offering it alone does not significantly raise the summary measure of engagement: the point estimate is 3.2 percentage points and indistinguishable from zero. Only the arm combining it with the peer conversation moves that measure, by 7.5 percentage points. In addition, the observed increase

in hospital visits suggests the presence of a salience or priming effect, whereby the intervention encourages individuals to seek care outside the programme's partner network.

The peer conversation does nothing on its own, yet the combined arm is the only one that clearly moves our measures of take-up, and the helpline offered alone moves them barely. One reason is that we do not measure everything: our outcomes are limited to administrative traces of contact with the health system, and do not directly measure stigma, which is precisely what the peer conversation targets. A refugee who leaves that conversation more willing to seek care leaves no trace unless the willingness passes through a channel we observe. The helpline is such a channel but a private decision to consult a general practitioner or to talk to a relative is not. In the latter case, the peer arm may do exactly what it was designed to do while remaining invisible in the data we currently hold, and what the combined arm adds is not motivation but the informational and administrative route that converts motivation into a recorded act. The protocol makes this plausible, since the helpline number is handed over and discussed during the same meeting, so the conversation can be what determines whether the number is later used.

However, this is not a limitation we are left to speculate about as separating the two channels is what the four-arm design was built for, and the endline survey will measure stigma and attitudes towards mental healthcare directly. Two outcomes are then possible, and they are informative in opposite ways. If the peer arm has reduced stigma without moving recorded care, then the barrier it targets is real but insufficient on its own, and information and navigation support are the complement that turns a change in attitudes into a change in behaviour. If stigma has not moved either, then a single conversation is a dose too small to shift it, which is what the intensity of comparable interventions would suggest: the visualisation treatment of John and Orkin (2022), which raises objectively measured water chlorination by 5 percentage points off an active-control mean of 23% and is still visible almost three years later, delivers two sessions of two hours a week apart, and the psychoeducation programme of Islam et al. (2024) delivers 44 weekly sessions over a year for reductions of 0.23 and 0.14 standard deviations in trauma and depressive symptoms.

Read against this literature, the magnitudes we report are not small. The combined arm raises the probability of any recorded contact by 7.5 percentage points on a control rate of 6.2%, more than doubling it, where the deadline-reminder letters of Domurat et al. (2021) raise marketplace enrolment by 1.3 percentage points on a control mean of 8.1%, a 16% increase, and the simplified notices of Bhargava and Manoli (2015a) raise EITC claiming from 0.14 to 0.23. What is small in our setting is the base rate and the absolute magnitude, not the proportional response, which is at the upper end of what light-touch interventions achieve. These estimates should moreover be read as a lower bound. They count only the contacts recorded by our partner structures, the public hospital system and the helpline, and miss care obtained anywhere else; the planned matching with national administrative data on healthcare consumption will capture the rest and deliver more precise estimates of the same effects.

6.2 Results on housing, employment and exit of the programme

We now turn to the second layer of our theory of change, namely whether the interventions translate into measurable progress on the employment and housing outcomes that constitute the core objectives of the AGIR programme. We exploit the manual collection of complete employment and housing trajectories described in B.1. For each refugee, we measure the share of days spent in a given state between the treatment date and the earlier of the AGIR exit and the measurement date, so that the denominator only covers the period over which the refugee is actually observed. Estimates follow the specification of Equation 3.2, with control for the pre-treatment measure of the outcome and with two measures of programme timing, thus exploiting the persistence of these trajectories to gain precision without affecting identification.

Employment. On the extensive margin, we consider the share of days in which the refugee is active, that is, in employment, education or training, and the share of days in employment only. Control refugees are active on 63.7% and employed on 58.3% of their observed days. No treatment arm displays a statistically significant effect on either outcome, and the joint test that the three arm coefficients are zero is never rejected. Point estimates are small and mostly negative, the largest being a 2.1 percentage point reduction in days employed for the peer-helper arm, which reaches 4.6 percentage points once baseline covariates are added but remains insignificant. On the intensive margin, which asks about the quality rather than the existence of employment, the picture is the same for full-time employment. Days in stable employment, that is, in open-ended or six-month-plus contracts, display borderline negative coefficients for the peer-helper and helpline arms, of about 5 percentage points against a control mean of 47.3%, but these do not survive the multiple-testing adjustment within the family and we do not read them as evidence of an effect. The full estimates are reported in Appendix D.3, Tables D6 and D7.

Housing. Control refugees spend 20.1% of their observed days in stable housing, and 9.8% in housing held in their own name, the most autonomous of the arrangements recorded. Neither margin responds to any of the three treatments, with point estimates of at most 3 percentage points in absolute value and joint tests never significant. Figure D1 displays the six employment and housing outcomes side by side, and no treated bar separates from its control mean.

Exit rates. Finally, we investigate whether the intervention accelerates refugees' exit from the AGIR programme prior to the mandatory 24-month administrative cap.[42] Figure 3.11 reports the probability that a refugee remains enrolled in the AGIR programme as a function of the number of days since the treatment date. To account for right-censoring induced by heterogeneous follow-up durations at the data cut-off, we report non-parametric Kaplan–Meier estimates of the survival function, *i.e.*, the probability of still being enrolled in the programme over time, separately for each treatment arm. In this framework, any recorded exit is treated as

[42] Table B1 reports the distribution of the 14 observed AGIR exit motives across the four treatment arms. We group these motives into four categories: *Positive* (exits associated with stable housing and/or stable employment), *Intermediate* (transitions to dedicated housing programmes or referrals to local service providers), *Administrative cap* (reaching the 24-month programme limit, recorded at the end of the month preceding the expected cap date), and *Negative/Other* (dropout, exclusion, emergency shelter, and residual categories).

an event, while refugees still enrolled at the cut-off date are right-censored at their individual follow-up duration. Overall retention does not differ across arms. The Kaplan-Meier curves largely overlap throughout the follow-up period, and even after 300 days since treatment no discernible difference emerges between groups. This is unsurprising: most recorded exits are administrative, and the intervention has no reason to move the date at which the 24-month ceiling is reached.

However, aggregate retention masks the composition of exits. We therefore decompose the observation window into four mutually exclusive and exhaustive states, so that each refugee is in exactly one of them at the measurement date and the four arm coefficients sum to zero by construction: (i) exited positively for permanent housing, for stable employment, or into a temporary accommodation centre for recognised refugees[43]; (ii) exited for another reason, pooling departmental moves, referrals, emergency accommodation, exclusions and dropouts; (iii) reached the 24-month administrative cap; or (iv) still enrolled. Table D5 reports the corresponding estimates. In the control group, 10.7% of refugees have exited positively, 4.1% for another reason, 39.0% have reached the cap and 46.2% are still enrolled.

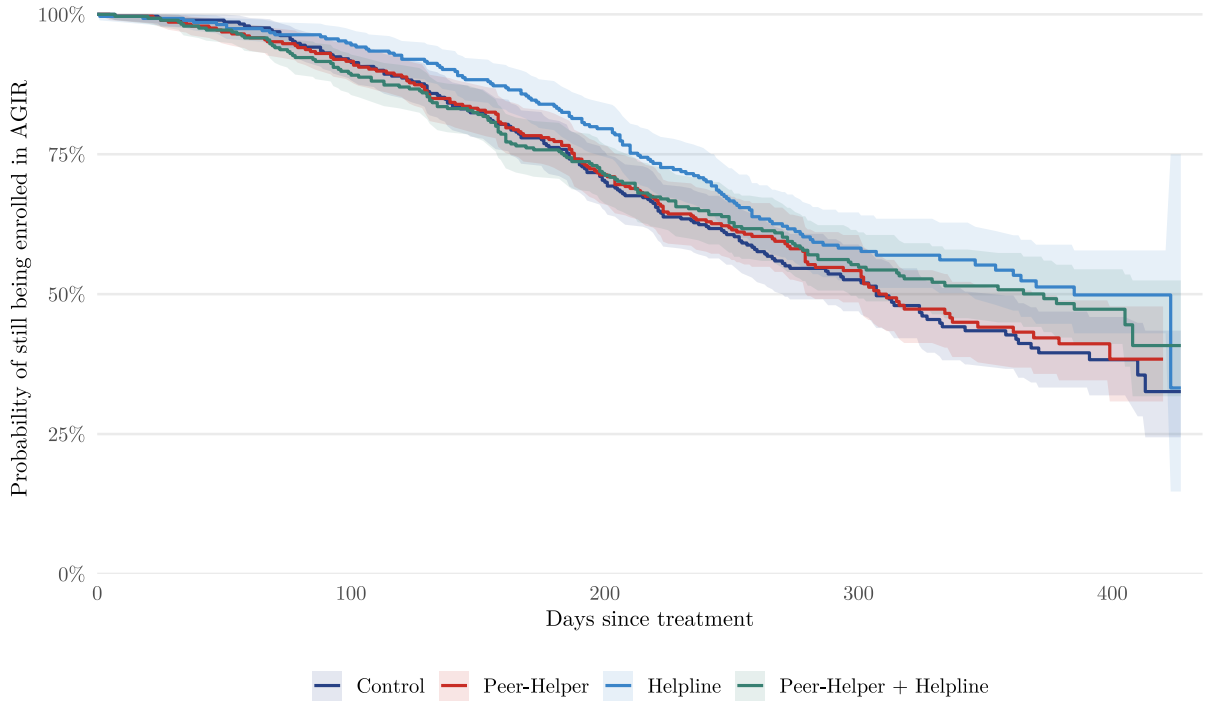

Figure 3.11: Probability of still being enrolled in AGIR, by treatment arm

Notes: Step curves are Kaplan-Meier estimates of the probability that a refugee is still enrolled in the AGIR programme as a function of the number of days since the treatment date. Any recorded exit (early exit or *Reached maximum support period*) counts as an event; refugees still enrolled at the AGIR monitoring cut-off (28 July 2026) are right-censored at their individual follow-up duration. Shaded bands are pointwise 95% confidence intervals. Because the maximum follow-up in this extraction is about 10 months while the AGIR administrative cap is at 24 months, the curves only begin to descend and will fill in as later cohorts reach the cap.

[43] CPH, *centre provisoire d'hébergement*, a dedicated follow-on housing scheme that caseworkers record as an exit only when it resolves the housing situation

Contrary to what the employment and housing results just documented would lead one to expect, two patterns stand out, and we are unable to account for either. First, assignment to the combined arm reduces the probability of a positive exit by 5.4 percentage points, that is, by half relative to the control group, and this reduction is matched almost exactly by a higher probability of still being enrolled at the measurement date. Adjusting for baseline characteristics brings the estimate to 4.5 percentage points, still significant at the 5% level. Other exits and cap exits are untouched, so this is not a substitution of positive exits towards worse ones: treated refugees are simply less likely to leave the programme with a housing or employment solution recorded, and correspondingly more likely to remain in it. Second, the helpline-only arm displays a different and equally puzzling pattern, with 7.4 fewer percentage points of cap exits and 10.0 more of refugees still enrolled; unlike the previous one, this shift is substantially attenuated once covariates are included, which points towards a chance imbalance in the timing of entry into AGIR rather than towards a behavioural response, since reaching the cap before the measurement date is largely determined by entry date.

Several mechanical explanations can be ruled out. The classification itself is sound: 94% of the refugees recorded as exiting positively are in stable housing on the exact day they leave the programme and 77% are employed, against 15% and 64% for those reaching the cap (Table B2). Composition is not the explanation either, although it accounts for part of the estimate. The covariate pool includes the refugee's housing and employment situation on the treatment day itself, which matters mechanically here since a positive exit requires a housing solution. The combined arm happens to draw slightly fewer refugees already in stable housing at randomisation, 11% against 14% in the control group, a gap that is not statistically significant (Table C1) but is large relative to an outcome whose control mean is 10.7%. Adjusting for it absorbs roughly a fifth of the coefficient, which remains significant at the 5% level, and the double-selection LASSO retains both indicators. The pool also includes the time elapsed between entry into AGIR and treatment, so the result is not driven by treated refugees having been enrolled for longer or shorter when they were treated. However, the estimate does not survive multiple-testing adjustment within the family, and the joint test that the three arm coefficients are zero is not rejected. Above all, we can document no channel: the trajectory results reported in Section 6.2 show that treated refugees obtain stable housing at the same rate and speed as control refugees, so they are not forgoing the housing that a positive exit records, and the intervention protocol keeps the mental-health programme strictly separate from the AGIR pathway, with mediators instructed not to discuss programme duration and caseworkers having no access to any information about participation or care. We therefore report this pattern as an unexplained regularity rather than as a finding. We stress that this is not an artefact of measuring outcomes at an interim date but longer follow-up will settle the form that effect takes. Support is capped at twenty-four months, so being still enrolled is a transitory state and every refugee eventually falls into one of the three exit categories. If assignment to the combined arm delays positive exits, the refugees still enrolled in that arm will leave positively at a higher rate than their control counterparts and the gap will close by the end of the support period; if it reduces them, those refugees will instead reach the cap and the gap will persist. Both are treatment effects, and we currently

have no mechanism for either.

The absence of any documented effect on housing and employment is what makes this result on the exits puzzling. Since treated refugees hold stable housing on the same share of days as control refugees, and reach it at the same rate and speed, the lower probability of a positive exit in the combined arm cannot be attributed to a housing solution that fails to materialise. What the intervention appears to change is whether an unchanged housing situation is converted into a recorded departure from the programme, not the housing situation itself. We do not have a mechanism that would produce such a wedge, and we therefore leave the question open.

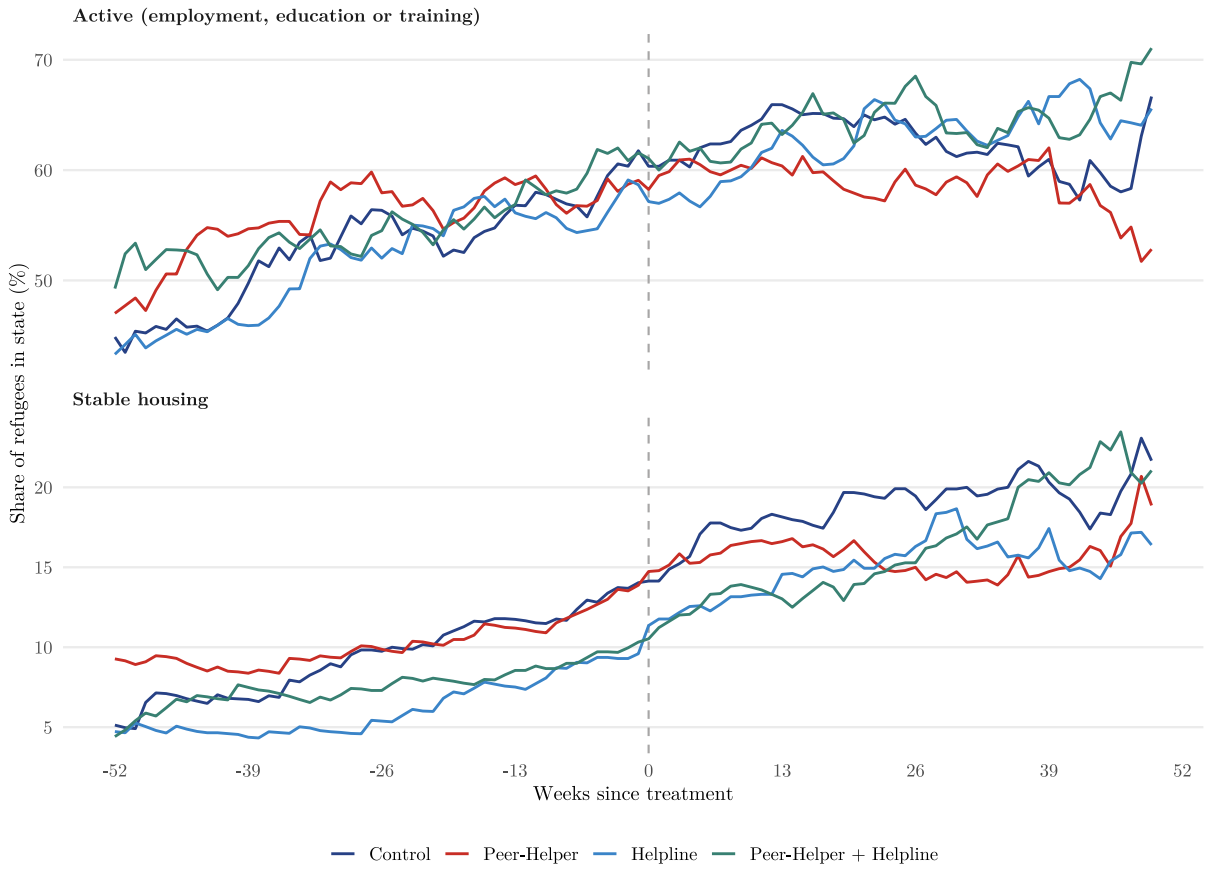

Figure 3.12: Share of refugees active and in stable housing, by week relative to treatment

Notes: Each curve reports, for one treatment arm, the share of refugees who are active (in employment, education or training, top panel) or in stable housing (bottom panel) in a given week relative to their individual treatment date, based on the manually collected employment and housing spells. The dashed vertical line marks the treatment date. A refugee enters the risk set at entry into the AGIR programme and leaves it at programme exit or at the measurement date; weeks in which any arm has fewer than 50 observable refugees are not displayed.

Figure 3.12 summarises these results in event time. The four arms follow each other closely both before and after the treatment date, with a common upward drift that reflects progress within the programme rather than any treatment-induced divergence. Two readings are consistent with this null. Given the modest effect of the interventions on healthcare engagement documented in the previous section, and the fact that the mechanism operates through mental health, which is itself a slow-moving input into employment and housing, the elapsed time may

simply be too short for any downstream effect to materialise. Alternatively, the pathway from mental healthcare to economic integration may be weaker than the theory of change posits, a question the endline survey and the longer follow-up are designed to settle.

Effects on the treated. Because the interventions move engagement with mental healthcare only modestly, instrumenting engagement by random assignment yields estimates that are too imprecise to be informative. With take-up between 3 and 5% the resulting confidence intervals for the LATE exclude only implausibly large effects. First-stage Kleibergen-Paap statistics range from 4.6, when engagement is measured by any post-treatment healthcare contact instrumented by the three assignment dummies, to 13.0 for the arm-specific helpline instrument. We report these estimates in Appendix D.5 for completeness, in Tables D13 and D14 for engagement instrumented by the three assignment dummies and Tables D15 and D16 for the arm-specific helpline instrument, and read them as a bounding exercise rather than as evidence on compliers. Table D17 applies the same instrument to the four exit states: it reproduces the sign pattern of the intention-to-treat estimates without adding precision. Beyond precision, the exclusion restriction is debatable here for the reasons set out in Section 5: assignment can affect employment and housing through channels other than the care contact we instrument, and both violations push towards overstating the complier effect. We therefore treat the two-stage estimates as an upper bound on what engagement with care could plausibly deliver, rather than as a causal estimate of it.

7 Heterogeneous effects

Gender. We estimate the benchmark specification with the three arm indicators fully interacted with a female indicator, so that the arm effect is identified separately for men and for women, the female main effect being absorbed by the randomisation strata. Women represent 199 of the 1,135 refugees in the analysis sample, which limits precision but not representativeness: because randomisation was stratified on *Groups of nationality* × *gender*, the female sub-sample is balanced across arms by construction, and the gender split is a dimension along which the design was explicitly powered to be informative. The full set of estimates is reported in E.2.

The main outcome family displaying an effect is hospital utilisation. The effect is entirely concentrated among women: assignment to the combined arm raises the probability of a hospital visit by 13 percentage points for women, against a control mean of 7.0%, while the corresponding estimate for men is 3 percentage points and not significant. The peer-helper arm alone moves women in the opposite direction, with a 7.6 percentage point reduction. Conversely, assignment to the combined arm has an effect on the helpline family outcomes but only for men with a 4 percentage points estimate for calling the helpline for psy reasons, 2.3 for securing an appointment, with almost perfect attendance. We are not able to estimate any effects for women. The summary any-signal outcome displays the same ordering, 16 percentage points for women against 6 for men, both significant at the 5% level. However, the gender difference is not statistically significant in either family of outcomes, and estimates do not survive the multiple hypothesis correction applied within each family. The three families are reported in Tables

E2, E5 and E1, and Table E3 adds the baseline covariates. No difference appears for partner-structure, exit or trajectory outcomes, reported in Tables E4, E6, E7 and E8. As discussed in Section 3, women are less likely to have been seen by the project managers in the first place and are disproportionately more educated so the heterogeneous effects may be identifying the characteristics that brought these women into the study rather than gender per say.

Peer-helpers. As set out in Section 5, the peer-helper intervention bundles a message with a messenger, and whether its effect travels with the content of the conversation or with the person delivering it determines whether investment should be put only with the message in the right language or if it is important to invest in the ability to recruit and train peers of comparable quality. Our design does not answer this by construction: assignment to the peer-helper arm is random, but the identity of the peer helper is not, depending on the participant's intervention language and gender and on which mediators are on duty that day.

We exploit the assignment schedules to reconstruct, for each control refugee, the peer helper they *would have* been assigned had they been allocated to a peer arm. We do this applying the same rules usually applied in the field by first checking the mediators available on the refugee's treatment half-day, then the language and gender priority lists. We validate this procedure by comparing it against the mediators actually assigned to treated refugees, whom it recovers exactly in 72.9% of cases; the procedure and its validation are detailed in E.2. Because the resulting mediator cell is a deterministic function of pre-treatment characteristics, assignment remains random within each cell, and the effect of peer assignment can be estimated against control refugees who would have faced the same peer. Table 3.1 reports these effects separately for the peer-helper-only and the combined arm.

Two patterns emerge. First, the effect of peer assignment is concentrated in the combined arm and absent from the peer-helper-only arm, in every mediator cell without exception. Second, within the combined arm, two mediators account for most of the effect, with increases in the probability of any post-treatment signal of 21.0 and 10.3 percentage points, the first of which survives the Benjamini-Hochberg correction, while the remaining cells are indistinguishable from zero.

Nonetheless, these estimates should be read as descriptive. Mediator cells are defined by language, gender and treatment date, so a cell confounds the mediator with the population they serve: a comparison across mediators is a comparison of peer-times-clientele combinations, not of mediator quality. Coefficients are therefore not directly comparable across cells, except among mediators who serve the same language and whose assignment is effectively driven by availability alone, which is the case of the three Pashto-speaking peers.

Table 3.1: Peer-assignment effect by (counterfactual) mediator and arm

	Any post-treatment signal		Any hospital visit				
Mediator	PH vs C	PH+HL vs C	PH vs C	PH+HL vs C	N Control	N PH	N PH+HL
PH French-Arabic	-0.017	0.210***	-0.045	0.102*	78	66	63
	(0.046)	(0.066)	(0.041)	(0.058)			
	[0.997]	[0.009]	[0.868]	[0.247]			
PH Pashto A	-0.000	-0.010	-0.001	-0.040	53	47	64
	(0.050)	(0.042)	(0.050)	(0.036)			
	[0.997]	[0.816]	[0.992]	[0.442]			
PH Pashto B	-0.039	0.068	-0.036	0.069	54	51	43
	(0.047)	(0.066)	(0.047)	(0.066)			
	[0.808]	[0.498]	[0.868]	[0.442]			
PH Bengali A	0.031	0.103**	0.032	0.082**	24	44	45
	(0.030)	(0.044)	(0.030)	(0.038)			
	[0.808]	[0.056]	[0.868]	[0.193]			
PH Tamil-English	-0.098	-0.095	-0.021	-0.023	14	35	27
	(0.104)	(0.105)	(0.081)	(0.085)			
	[0.808]	[0.498]	[0.992]	[0.788]			
Other mediator	0.005	0.045	-0.002	0.040	67	43	43
	(0.046)	(0.056)	(0.038)	(0.050)			
	[0.997]	[0.498]	[0.992]	[0.497]			

Notes: Each cell of the table reports the effect of assignment to one peer arm (Peer-Helper alone, or Peer-Helper + Helpline) versus Control within one mediator cell, from a single OLS regression per outcome of the outcome on mediator-cell indicators and their interactions with the two arm indicators, with randomisation strata fixed effects (site x nationality group x gender) and HC1 heteroskedasticity-robust standard errors in parentheses; Benjamini-Hochberg q-values in square brackets, pooled across the six mediator cells of each column. The mediator cell is the actual mediator for peer-arm refugees and the counterfactual mediator reconstructed by the assignment algorithm for Control refugees. Anonymised labels list the mediator's main intervention languages, with letters ranking mediators of the same language set by caseload; the Pashto mediators also serve Dari and Urdu, and the Bengali mediators Hindi. Cells are defined by language, gender and treatment date, so differences across mediators mix mediator quality with the composition of the population served. * p<0.10, ** p<0.05, *** p<0.01.

Finally, the joint test that the effect is equal across cells is not rejected, so the contrast between mediators is suggestive rather than established. With these reservations, the evidence speaks more to the message than to the messenger: the peer-helper arm is null in every cell, including those where the combined arm produces its largest effects, which is difficult to reconcile with a story in which a one-off peer conversation fails only because it was delivered by the wrong peers.

8 Conclusion

This paper asks whether low-cost, demand-side interventions can raise take-up of mental healthcare in a setting where the barriers are not only informational but also socio-cultural. That combination is what makes refugee mental healthcare a demanding test. The literature on take-up has established that light-touch interventions move behaviour when the binding constraint is information or administrative complexity, and it has established that stigma matters, but it has

rarely been able to say which of the two binds, because most designs vary only one at a time. Our four-arm experiment varies them separately and jointly: a confidential conversation with a trained former refugee in the participant's own language, a multilingual helpline that books appointments on the refugee's behalf, and their combination.

Our preliminary results show that assignment to the combined arm raises the probability of any recorded health or mental health-related signal by 7.5 percentage points against a control rate of 6.2%, an effect that survives adjustment for multiple hypotheses. Most of this response takes the form of hospital admissions rather than appointments with the programme's partner structures: the intervention appears to make care salient, and refugees then turn to the mainstream system. The helpline offered alone raises the probability of a call by 3.0 percentage points but moves the summary measure of signal only by an insignificant 3.2 points, and the peer conversation on its own moves nothing we observe. In absolute terms these are small numbers, on a small base; in proportional terms the combined arm more than doubles engagement, which is at the upper end of what light-touch interventions in this literature achieve. They are also a lower bound, since we currently observe only care obtained through our partner structures, the public hospital system and the helpline. Interestingly, we find suggestive evidence that these effects seem bigger when targeting women.

However, this design cannot yet say whether the peer conversation reduces stigma, because stigma is not among the outcomes our administrative data record. It can say that whatever the conversation does, it does not by itself produce a recorded contact with care, but that adding the helpline does. The endline survey, which measures attitudes towards mental healthcare directly, will separate the two readings that remain open: a fall in stigma without a rise in recorded care would show that the socio-cultural barrier is real but not sufficient on its own, while an unchanged stigma would indicate that a single conversation is too small a dose to shift it.

Two further steps are required to fully answer these questions. Matching with national administrative data on healthcare consumption, social benefits and labour-market histories will capture the care we currently miss and extend the follow-up well beyond the present window. And the Essonne site, whose intervention ended in June 2026, will increase the sample once its follow-up is comparable to Paris. Until then, we report evidence on the first layer of the theory of change only: refugees can be brought into contact with formal mental healthcare, modestly, and only when the two barriers are addressed together. Whether that contact improves mental health, and whether better mental health supports the housing and employment outcomes that the AGIR programme exists to deliver, remains to be established as we do not find evidence on longer-term integration outcomes.

References

Abel, Martin, Rulof Burger, Eliana Carranza, and Patrizio Piraino (2019). "Bridging the Intention-Behavior Gap? The Effect of Plan-Making Prompts on Job Search and Employment". *American Economic Journal: Applied Economics* 11.2, pp. 284–301 (cit. on p. 171).

Abramitzky, Ran, Leah Boustan, Peter Catron, Dylan Connor, and Rob Voigt (2023). "The refugee advantage: English-language attainment in the early twentieth century". *Sociological Science* 10, pp. 769–805 (cit. on p. 174).

Adda, Jérôme, Christian Dustmann, and Joseph-Simon Görlach (2022). "The Dynamics of Return Migration, Human Capital Accumulation, and Wage Assimilation". *The Review of Economic Studies* 89.6, pp. 2841–2871. eprint: https://academic.oup.com/restud/article-pdf/89/6/2841/46869661/rdac003.pdf (cit. on p. 174).

Aksoy, Cevat Giray, Gaurav Khanna, Victoria Marino, and Semih Tumen (2024). "Hometown Conflict and Refugees' Integration Efforts". *Available at SSRN 4758122* (cit. on p. 174).

Andersson Joona, Pernilla, Alma W Lanninger, and Marianne Sundström (2016). *Reforming the integration of refugees: The Swedish experience* (cit. on p. 174).

Bahar, Dany, Rebecca J Brough, and Giovanni Peri (2024). *Forced Migration and Refugees: Policies for Successful Economic and Social Integration* (cit. on p. 174).

Banerjee, Souvik, Pinka Chatterji, and Kajal Lahiri (2017). "Effects of psychiatric disorders on labor market outcomes: A latent variable approach using multiple clinical indicators". *Health economics* 26.2, pp. 184–205 (cit. on p. 171).

Barenbaum, Joshua, Vladislav Ruchkin, and Mary Schwab-Stone (2004). "The psychosocial aspects of children exposed to war: Practice and policy initiatives". *Journal of child psychology and psychiatry* 45.1, pp. 41–62 (cit. on p. 171).

Batista, Catia, Lara Bohnet, Jules Gazeaud, and Julia Seither (2025). "Integrating Migrants: Experimental Evidence on Cross-Border Spillovers" (cit. on p. 172).

Battisti, Michele, Yvonne Giesing, and Nadzeya Laurentsyeva (2019). "Can job search assistance improve the labour market integration of refugees? Evidence from a field experiment". *Labour Economics* 61, p. 101745 (cit. on pp. 172, 174).

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen (2014). "Inference on treatment effects after selection among high-dimensional controls". *Review of Economic Studies* 81.2, pp. 608–650 (cit. on p. 202).

Bhargava, Saurabh and Dayanand Manoli (2015a). "Psychological Frictions and the Incomplete Take-Up of Social Benefits: Evidence from an IRS Field Experiment". *The American Economic Review* 105.11, pp. 3489–3529 (cit. on pp. 171, 203).

— (2015b). "Psychological frictions and the incomplete take-up of social benefits: Evidence from an IRS field experiment". *American Economic Review* 105.11, pp. 3489–3529 (cit. on p. 174).

Blackmore, Rebecca, Jacqueline A Boyle, Mina Fazel, Sanjeeva Ranasinha, Kylie M Gray, Grace Fitzgerald, et al. (2020). "The prevalence of mental illness in refugees and asylum seekers: A systematic review and meta-analysis". *PLoS medicine* 17.9 (cit. on p. 171).

Bogic, Marija, Anthony Njoku, and Stefan Priebe (2015). "Long-term mental health of war-refugees: A systematic literature review". *BMC international health and human rights* 15, pp. 1–41 (cit. on p. 171).

Braun, Sebastian Till and Nadja Dwenger (2019). *Settlement Location Shapes Refugee Integration: Evidence from Post-War Germany* (cit. on p. 174).

Brell, Courtney, Christian Dustmann, and Ian Preston (2020). "The labor market integration of refugee migrants in high-income countries". *Journal of Economic Perspectives* 34.1, pp. 94–121 (cit. on p. 171).

Brodeur, Abel (2025). *Status on GDRI And Closely Related Studies* (cit. on p. 174).

Brodeur, Abel, Lenka Fiala, Jack Fitzgerald, Essi Kujansuu, David Valenta, Ole Rogeberg, et al. (2025). *A Comment on "Improving Women's Mental Health During a Pandemic"* (cit. on p. 174).

Brouwer, Thijs, Fabio Galeotti, and Marie Claire Villeval (2023). "Teaching Norms: Direct Evidence of Parental Transmission". *The Economic Journal* 133.650, pp. 872–887. eprint: https://academic.oup.com/ej/article-pdf/133/650/872/51839901/ueac074.pdf (cit. on p. 174).

Brücker, Herbert, Nina Rother, Jürgen Schupp, et al. (2018). *IAB-BAMF-SOEP-Befragung von Geflüchteten 2016: Studiendesign, Feldergebnisse sowie Analysen zu schulischer wie beruflicher Qualifikation, Sprachkenntnissen sowie kognitiven Potenzialen*. Vol. 30. DEU (cit. on p. 186).

Bryant, Richard A, Angela Nickerson, Naser Morina, and Belinda Liddell (2023). "Posttraumatic stress disorder in refugees". *Annual review of clinical psychology* 19.1, pp. 413–436 (cit. on p. 171).

Burnett, Angela and Michael Peel (2001). "The health of survivors of torture and organised violence". *Bmj* 322.7286, pp. 606–609 (cit. on p. 171).

Byrow, Yulisha, Rosanna Pajak, Tadgh McMahon, Amitabh Rajouria, and Angela Nickerson (2019). "Barriers to mental health help-seeking amongst refugee men". *International journal of environmental research and public health* 16.15, p. 2634 (cit. on p. 171).

Celhay, Pablo, Bruce D Meyer, and Nikolas Mittag (2025). "Stigma in welfare programs". *Review of Economics and Statistics*, pp. 1–37 (cit. on p. 174).

Chatterji, Pinka, Margarita Alegría, Mingshan Lu, and David Takeuchi (2007). "Psychiatric disorders and labor market outcomes: Evidence from the National Latino and Asian American Study". *Health Economics* 16.10, pp. 1069–1090. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hec.1210 (cit. on p. 171).

Cloitre, M., M. Shevlin, C. R. Brewin, J. I. Bisson, N. P. Roberts, A. Maercker, et al. (2018). "The International Trauma Questionnaire: Development of a self-report measure of ICD-11 PTSD and complex PTSD". *Acta Psychiatrica Scandinavica* 138.6, pp. 536–546. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/acps.12956 (cit. on pp. 185, 231).

Couttenier, Mathieu, Veronica Petrencu, Dominic Rohner, and Mathias Thoenig (2019). "The Violent Legacy of Conflict: Evidence on Asylum Seekers, Crime, and Public Policy in Switzerland". *American Economic Review* 109.12, pp. 4378–4425 (cit. on pp. 171, 174).

Currie, Janet (2004). *The take up of social benefits* (cit. on p. 171).

Dang, Hai-Anh H., Trong-Anh Trinh, and Paolo Verme (2023). "Do refugees with better mental health better integrate? Evidence from the Building a New Life in Australia longitudinal survey". *Health Economics* 32.12, pp. 2819–2835. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hec.4750 (cit. on pp. 171, 174).

Derluyn, Ilse, Eric Broekaert, Gilberte Schuyten, and Els De Temmerman (2004). "Post-traumatic stress in former Ugandan child soldiers". *The Lancet* 363.9412, pp. 861–863 (cit. on p. 171).

DeSa, Sarah, Akalewold T Gebremeskel, Olumuyiwa Omonaiye, and Sanni Yaya (2022). "Barriers and facilitators to access mental health services among refugee women in high-income countries: A systematic review". *Systematic reviews* 11.1, p. 62 (cit. on p. 171).

Di Thiene, D, Magnus Helgesson, S Rahman, K Alexanderson, J Tiihonen, G La Torre, et al. (2021). "Risk of labour market marginalisation among young refugees and non-refugee migrants with common mental disorders". *Social psychiatry and psychiatric epidemiology* 56, pp. 1025–1034 (cit. on p. 171).

Dietrich, Hans, José Luis Álvaro Estramiana, Alicia Garrido Luque, and Volker Reissner (2023). "Effects of Posttraumatic Stress Disorder and Mental Disorders on the Labor Market Integration of Young Syrian Refugees". *International Journal of Environmental Research and Public Health* 20.3, p. 2468 (cit. on p. 171).

Domurat, Richard, Isaac Menashe, and Wesley Yin (2021). "The role of behavioral frictions in health insurance marketplace enrollment and risk: Evidence from a field experiment". *American Economic Review* 111.5, pp. 1549–1574 (cit. on pp. 171, 203).

DSED (2026a). *Le programme d'accompagnement global et individualisé des réfugiés - AGIR Les entrées dans le programme* (cit. on pp. 177, 217).

— (2026b). *Le programme d'accompagnement global et individualisé des réfugiés - AGIR Les sorties du programme* (cit. on p. 177).

Dustmann, Christian, Francesco Fasani, Tommaso Frattini, Luigi Minale, and Uta Schönberg (2017). "On the economics and politics of refugee migration". *Economic policy* 32.91, pp. 497–550 (cit. on pp. 171, 174).

Dustmann, Christian and Joseph-Simon Görlach (2016). "The economics of temporary migrations". *Journal of Economic Literature* 54.1, pp. 98–136 (cit. on p. 174).

Dustmann, Christian, Rasmus Landersø, and Lars Højsgaard Andersen (2024). "Refugee benefit cuts". *American Economic Journal: Economic Policy* 16.2, pp. 406–441 (cit. on p. 174).

ELIPA2 (2024a). *La santé générale et mentale des nouveaux migrants* (cit. on p. 191).

— (2024b). *La seconde édition de l'Enquête Longitudinale sur l'Intégration des Primo-Arrivants* (cit. on p. 175).

— (2024c). *Le parcours d'intégration des réfugiés* (cit. on p. 176).

Fasani, Francesco, Tommaso Frattini, and Luigi Minale (2021a). "(The Struggle for) Refugee integration into the labour market: Evidence from Europe". *Journal of Economic Geography* 22.2, pp. 351–393. eprint: https://academic.oup.com/joeg/article-pdf/22/2/351/42902602/lbab011.pdf (cit. on p. 171).

— (2021b). "Lift the ban? Initial employment restrictions and refugee labour market outcomes". *Journal of the European Economic Association* 19.5, pp. 2803–2854 (cit. on p. 174).

Fazel, Mina, Jeremy Wheeler, and John Danesh (2005). "Prevalence of serious mental disorder in 7000 refugees resettled in western countries: A systematic review". *The Lancet* 365.9467, pp. 1309–1314 (cit. on p. 171).

Foged, Mette, Linea Hasager, and Giovanni Peri (2024a). "Comparing the Effects of Policies for the Labor Market Integration of Refugees". *Journal of Labor Economics* 42.S1, S335–S377 (cit. on p. 174).

Foged, Mette, Linea Hasager, Giovanni Peri, Jacob Nielsen Arendt, and Iben Bolvig (2024b). "Language Training and Refugees' Integration". *Review of Economics and Statistics* 106.4, pp. 1157–1166 (cit. on p. 174).

Foged, Mette, Karen-Inge Karstoft, and Edith Zink (2026). *PTSD and Refugees' Underemployment: Evidence from Displaced Ukrainians*. Discussion Paper Series 019/26. ROCKWOOL Foundation Berlin (RFBerlin) (cit. on pp. 171, 174, 190).

Fouka, Vasiliki (2023). "State policy and immigrant integration". *Annual Review of Political Science* 27 (cit. on p. 174).

Frijters, Paul, David W. Johnston, and Michael A. Shields (2010). *Mental Health and Labour Market Participation: Evidence from IV Panel Data Models* (cit. on p. 171).

Fundamental Rights, European Union Agency for (2011). *Fundamental rights of migrants in an irregular situation in the European Union* (cit. on p. 171).

Giuntella, Osea, Zovanga L Kone, Isabel Ruiz, and Carlos Vargas-Silva (2018). "Reason for immigration and immigrants' health". *Public health* 158, pp. 102–109 (cit. on p. 171).

Gleeson, Christina, Rachel Frost, Larissa Sherwood, Mark Shevlin, Philip Hyland, Rory Halpin, et al. (2020). "Post-migration factors and mental health outcomes in asylum-seeking and refugee populations: A systematic review". *European Journal of Psychotraumatology* 11.1, p. 1793567 (cit. on p. 171).

Henderson, Claire and Graham Thornicroft (2009). "Stigma and discrimination in mental illness: Time to Change". *The Lancet* 373.9679, pp. 1928–1930 (cit. on p. 174).

Islam, Asad, Tanvir Ahmed Mozumder, Tabassum Rahman, Tanvir Shatil, and Abu Siddique (2024). *Forced Displacement, Mental Health, and Child Development: Evidence from the Rohingya Refugees*. Tech. rep. IZA Discussion Papers (cit. on pp. 173, 174, 203).

John, Anett and Kate Orkin (2022). "Can simple psychological interventions increase preventive health investment?" *Journal of the European Economic Association* 20.3, pp. 1001–1047 (cit. on pp. 171, 203).

Johnson, Howard and Andrew Thompson (2008). "The development and maintenance of post-traumatic stress disorder (PTSD) in civilian adult survivors of war trauma and torture: A review". *Clinical psychology review* 28.1, pp. 36–47 (cit. on p. 171).

Kiselev, Nikolai, Monique Pfaltz, Florence Haas, Matthis Schick, Marie Kappen, Marit Sijbrandij, et al. (2020). "Structural and socio-cultural barriers to accessing mental healthcare among Syrian refugees and asylum seekers in Switzerland". *European journal of psychotraumatology* 11.1, pp. 1717–825 (cit. on p. 172).

Ko, Wonsik and Robert A Moffitt (2022). "Take-up of social benefits". *NBER Working Paper* w30148 (cit. on p. 171).

Kroenke, Kurt, Robert L. Spitzer, and Janet B. W. Williams (2003). "The Patient Health Questionnaire-2: Validity of a Two-Item Depression Screener". *Medical Care* 41.11, pp. 1284–1292 (cit. on pp. 189, 231, 233).

Langlois, Etienne V, Andy Haines, Göran Tomson, and Abdul Ghaffar (2016). "Refugees: towards better access to health-care services". *Lancet (London, England)* 387.100-16, p. 319 (cit. on p. 171).

Liu, Xiangqing, Elisabetta Lodigiani, Silvana Robone, Elisa Tosetti, and Giorgio Vittadini (2025). *The Effect of the Great Recession on the Mental Health Care of Immigrant and Native Workers: Evidence from Administrative Data in Italy* (cit. on pp. 171, 174).

Lochmann, A, H Rapoport, and B Speciale (2019). "The effect of language training on immigrants' economic integration: Empirical evidence from France ". *European Economic Review* 113, pp. 265–296 (cit. on p. 174).

Mellor, Rachel, Allison Werner, Batool Moussa, Mohammed Mohsin, Rohan Jayasuriya, and Alvin Kuowei Tay (2021). "Prevalence, predictors and associations of complex post-traumatic stress disorder with common mental disorders in refugees and forcibly displaced populations: A systematic review". *European journal of psychotraumatology* 12.1, p. 1863579 (cit. on p. 171).

Moya, Andrés, Maria José Torres, Juliana Sánchez-Ariza, Arturo Harker, Alicia Lieberman, Blasina Nino, et al. (2024). "Caregiver Mental Health and Early Childhood Development: Experimental Evidence from a Conflict-Affected Setting" (cit. on pp. 173, 174).

Niederkrotenthaler, Thomas, Ellenor Mittendorfer-Rutz, Fredrik Saboonchi, and Magnus Helgesson (2020). "The role of refugee status and mental disorders regarding subsequent labour market marginalisation: a register study from Sweden". *Social psychiatry and psychiatric epidemiology* 55, pp. 697–704 (cit. on p. 171).

Norredam, Marie, Anna Mygind, and Allan Krasnik (2005). "Access to health care for asylum seekers in the European Union comparative study of country policies". *European Journal of Public Health* 16.3, pp. 285–289. eprint: https://academic.oup.com/eurpub/article-pdf/16/3/285/1272432/cki191.pdf (cit. on p. 171).

OFII (2024). *Régulation des orientations vers le programme AGIR* (cit. on p. 177).

OFPRA (2024). *Office français de protection des réfugiés et apatrides. Rapport d'activité 2023* (cit. on p. 175).

— (2025). *Office français de protection des réfugiés et apatrides. Rapport d'activité 2024* (cit. on pp. 171, 175).

Ojeda, Victoria D, Richard G Frank, Thomas G McGuire, and Todd P Gilmer (2010). "Mental illness, nativity, gender and labor supply". *Health Economics* 19.4, pp. 396–421 (cit. on p. 171).

Ridley, Matthew, Gautam Rao, Frank Schilbach, and Vikram Patel (2020). "Poverty, depression, and anxiety: Causal evidence and mechanisms". *Science* 370.6522, eaay0214 (cit. on p. 171).

Romano, Joseph P and Michael Wolf (2005). "Stepwise multiple testing as formalized data snooping". *Econometrica* 73.4, pp. 1237–1282 (cit. on p. 193).

Rosholm, Michael and Rune Vejlin (2010). "Reducing income transfers to refugee immigrants: Does start-help help you start?" *Labour Economics* 17.1, pp. 258–275 (cit. on p. 174).

Rozo, Sandra V and Guy Grossman (2025). *Refugees and Other Forcibly Displaced Populations*. World Bank (cit. on p. 172).

Ruiz, Isabel and Carlos Vargas-Silva (2018). "Differences in labour market outcomes between natives, refugees and other migrants in the UK". *Journal of Economic Geography* 18.4, pp. 855–885. eprint: https://academic.oup.com/joeg/article-pdf/18/4/855/25142517/lby027.pdf (cit. on p. 171).

Sabety, Adrienne, Jonathan Gruber, Jin Yung Bae, and Rishi Sood (2023). "Reducing Frictions in Health Care Access: The ActionHealthNYC Experiment for Undocumented Immigrants". *American Economic Review: Insights* 5.3, pp. 327–46 (cit. on pp. 173, 179).

Salami, Bukola, Jordana Salma, and Kathleen Hegadoren (2019). "Access and utilization of mental health services for immigrants and refugees: Perspectives of immigrant service providers". *International Journal of Mental Health Nursing* 28.1, pp. 152–161 (cit. on p. 171).

Satinsky, Emily, Daniela C Fuhr, Aniek Woodward, Egbert Sondorp, and Bayard Roberts (2019). "Mental health care utilisation and access among refugees and asylum seekers in Europe: A systematic review". *Health Policy* 123.9, pp. 851–863 (cit. on p. 171).

Seiler, Natalie, Keivan Davoodi, Michael Keem, and Subhash Das (2023). "Assessment tools for complex post traumatic stress disorder: A systematic review". *International journal of psychiatry in clinical practice* 27.3, pp. 292–300 (cit. on p. 185).

Shevlin, Mark, Philip Hyland, Stephen Butter, Orla McBride, Todd K. Hartman, Thanos Karatzias, et al. (2023). "The development and initial validation of self-report measures of ICD-11 depressive episode and generalized anxiety disorder: The International Depression Questionnaire (IDQ) and the International Anxiety Questionnaire (IAQ)". *Journal of Clinical Psychology* 79.3, pp. 854–870 (cit. on pp. 231, 233).

Smith, Emma C. (2025). *Stigma and Social Cover: A Mental Health Care Experiment in Refugee Networks* (cit. on pp. 173, 174, 179, 180).

Steel, Zachary, Derrick Silove, Kevin Bird, Patrick McGorry, and P Mohan (1999). "Pathways from war trauma to posttraumatic stress symptoms among Tamil asylum seekers, refugees, and immigrants". *Journal of traumatic stress* 12, pp. 421–435 (cit. on p. 171).

Stillman, Sarah, Sandra V Rozo, Abdulrazzak Tamim, I Bailey Palmer, Emma Smith, and Edward Miguel (2022). "The Syrian refugee life study: First glance". *Oxford Review of Economic Policy* 38.3, pp. 625–653 (cit. on p. 171).

Teodorescu, Dinu-Stefan, Johan Siqveland, Trond Heir, Edvard Hauff, Tore Wentzel-Larsen, and Lars Lien (2012). "Posttraumatic growth, depressive symptoms, posttraumatic stress symptoms, post-migration stressors and quality of life in multi-traumatized psychiatric outpatients with a refugee background in Norway". *Health and quality of life outcomes* 10, pp. 1–16 (cit. on p. 171).

Tinghög, Petter, Andreas Malm, Charlotta Arwidson, Erika Sigvardsdotter, Andreas Lundin, and Fredrik Saboonchi (2017). "Prevalence of mental ill health, traumas and postmigration stress among refugees from Syria resettled in Sweden after 2011: A population-based survey". *BMJ open* 7.12, e018899 (cit. on p. 171).

Turner, Stuart W, Cameron Bowie, Graham Dunn, Laidon Shapo, and William Yule (2003). "Mental health of Kosovan Albanian refugees in the UK". *The British Journal of Psychiatry* 182.5, pp. 444–448 (cit. on p. 171).

UNHCR (2025). *UNHCR, Global Appeal 2025*. Tech. rep. (cit. on p. 171).

Urbina, Maria José, Sandra V. Rozo, Andrés Moya, and Ana María Ibáñez (2023). "Least Protected, Most Affected: Impacts of Migration Regularization Programs on Pandemic Resilience". *AEA Papers and Proceedings* 113, pp. 347–51 (cit. on p. 174).

Yang, Dean, James Allen Iv, Arlete Mahumane, James Riddell Iv, and Hang Yu (2023). "Knowledge, stigma, and HIV testing: An analysis of a widespread HIV/AIDS program". *Journal of Development Economics* 160, pp. 1029–58 (cit. on p. 174).

Yu, Hang (2023). "Social stigma as a barrier to HIV testing: Evidence from a randomized experiment in Mozambique". *Journal of Development Economics* 161, pp. 1030–35 (cit. on p. 174).

Appendix

A Implementation of the Intervention

A.1 Intervention materials

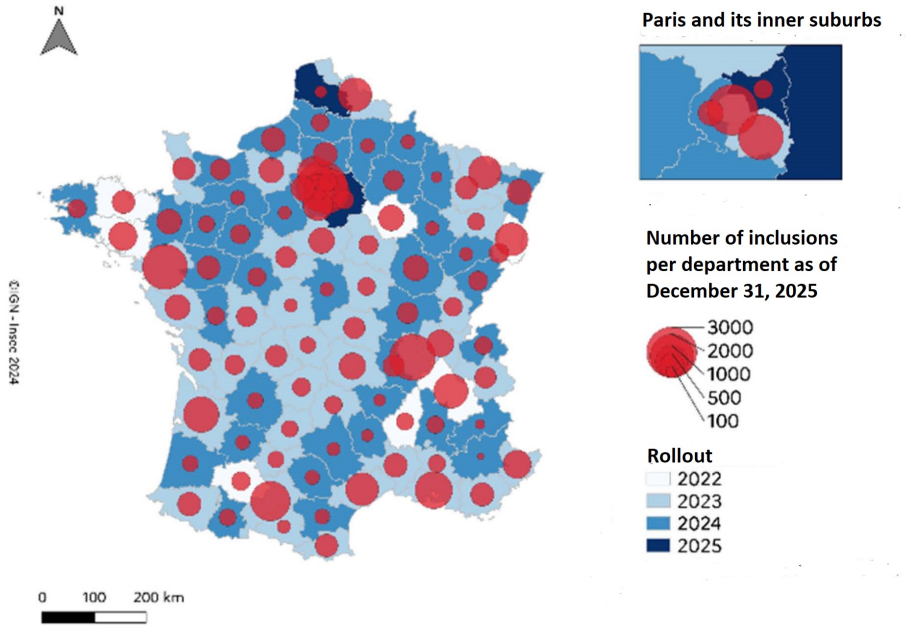

Figure A1: Rollout and inclusions in the AGIR programme
Sources: DSED, 2026a.

RISE

Figure A2: RISE Project Logo
Sources: Authors' elaboration.



Figure A3: Information Leaflet
Sources: Authors' elaboration.



Figure A4: Helpline Label
Sources: Authors' elaboration.



Figure A5: Video presentation of the project
Notes: The full English video is available upon request to the authors.
Sources: Authors' elaboration.

Video transcript:

Hello, I'm Alex from the RISE project. If you're watching this video, it's because we are inviting you to take part in an important initiative aimed at improving refugees' access to healthcare services. We're going to explain why we are carrying out this project, and what it means for you if you choose to participate. Our RISE team is focused on the experiences of refugees benefiting from the AGIR programme, the programme through which you are being supported here at France Terre d'Asile (COALLIA). France Terre d'Asile (COALLIA) helps you find housing and professional training. Our RISE team, however, is only focused on health. The RISE project takes place within the facilities of France Terre d'Asile (COALLIA) and is open to all refugees supported by AGIR in Paris and the Île-de-France region. However, it is a completely separate programme from AGIR. If you decide not to take part in our program, it will have no impact on your participation in AGIR. Our goal is simple: to better understand how we can help you access healthcare, especially mental health care. To do this, we want to understand your needs. We want to identify the challenges you may face and create practical solutions to improve healthcare access for all refugees. The RISE team is made up of researchers from universities and healthcare professionals, who ensure that the project remains confidential. This means that no personal data will be shared with anyone outside the RISE team. What will happen if you agree to participate? A member of our team will have a brief conversation with you about the project and your participation. Then, you'll be asked to answer a very short questionnaire about your health. This questionnaire is completely confidential and is only seen by the RISE project team. For example, France Terre d'Asile (COALLIA) will not have access to your answers. You will immediately receive an information guide in your language, with the addresses of healthcare centres and associations that offer care to people in exile. It's also possible that you will be offered an individual session today with a mediator who speaks your language, to give you more information about how to access care. We will also contact you again in a few months to collect a few quick updates about your health and healthcare needs. If you have any questions, feel free to ask the project staff who are here on-site. They'll be happy to welcome you and answer any questions you may have. Your experience is extremely valuable. By participating in this project, you are not only helping researchers and healthcare workers better understand the needs of refugees who face difficulties accessing care, but we also hope that your participation will help you receive better support for your own health needs. Thank you for watching this video. We hope to count you among our participants and wish you a wonderful day.

A.2 Sample construction and selection into the study

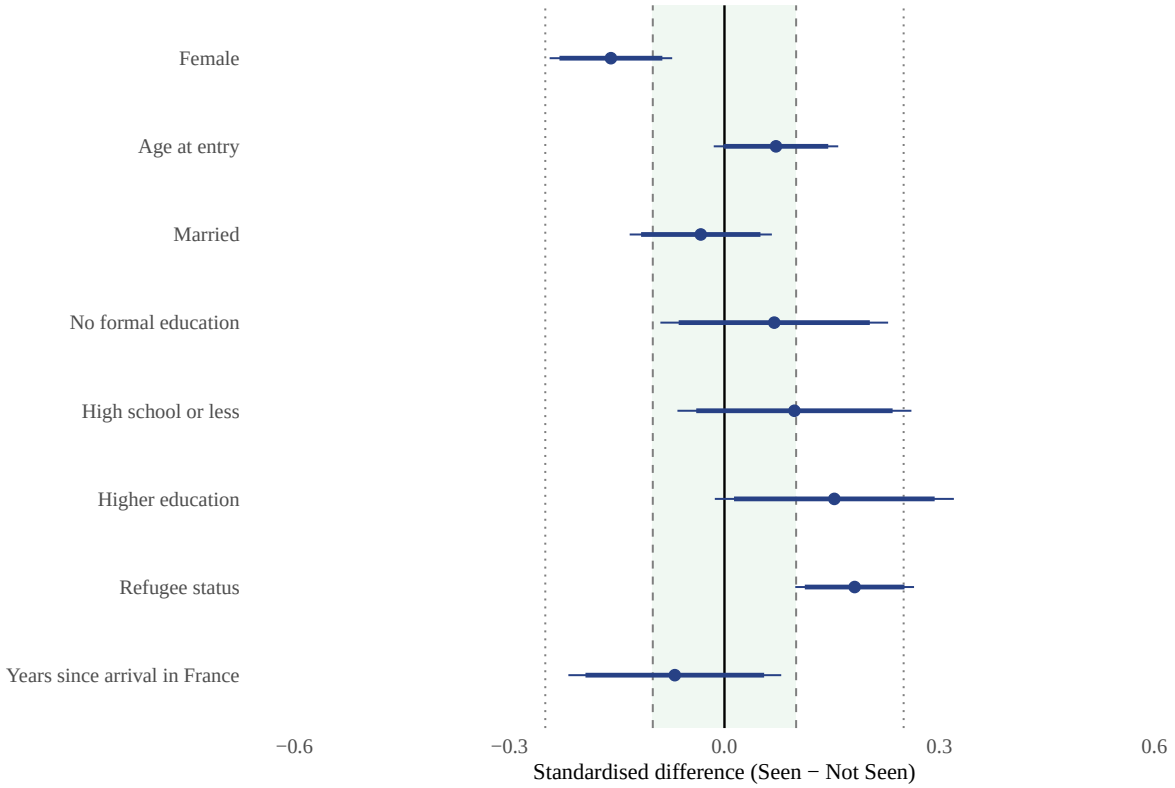

Figure A6: Covariate Balance Between Individuals Seen and Not Seen by RISE

Notes: For each baseline characteristic, the figure plots the standardised difference between individuals seen and not seen by RISE, computed as the seen-minus-not-seen mean divided by the not-seen-group standard deviation. Thick bars are 90% confidence intervals; thin bars are 95% confidence intervals. The inner dashed lines and shaded band delimit the $[-0.1, 0.1]$ range, within which standardised differences are typically considered negligible. The outer dotted lines delimit the $[-0.25, 0.25]$ range. "Seen by RISE" refers to individuals who were physically met by a RISE team member during a visit with their AGIR social worker, regardless of whether they subsequently consented to participate in the study. *Years since arrival in France*: difference in years between France arrival date and AGIR programme entry date. Observations with missing values are dropped on a per-variable basis. We note that the variables used in this figure are those extracted automatically from AGIR's information system. In the dataset used for analysis containing consenting individuals only, these variables have been corrected and supplemented, when missing, through manual data collection.

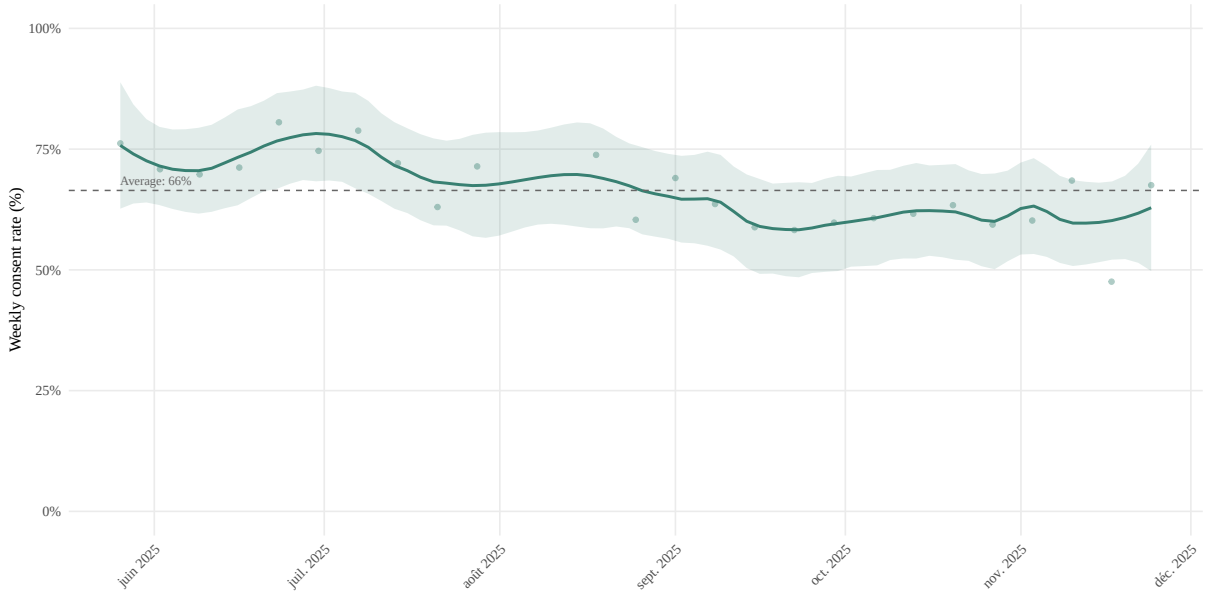

Figure A7: Consent Rate Over Time

Notes: Each dot represents the weekly consent rate (number of individuals who consented divided by the number seen). The solid line represents the smoothed fit of the weekly consent rate. The dashed line indicates the overall average consent rate of 66%. This figure concerns FTDA only, as COALLIA data collection is still ongoing.

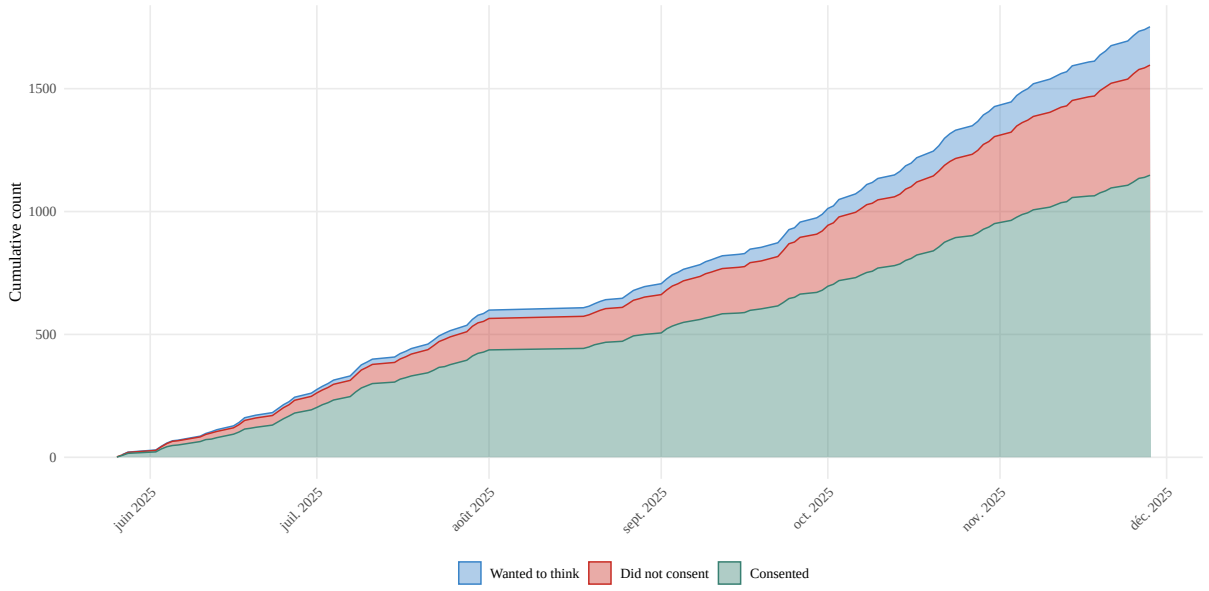

Figure A8: Cumulative Inclusion of Refugees in RISE

Notes: This figure displays the cumulative number of refugees approached by the project over the inclusion period (May–December 2025), by consent status. Areas are stacked, so the total height at any point represents the cumulative number of individuals seen. The green area represents individuals who consented, the red area those who did not consent, and the blue area those who requested additional time to decide. This figure concerns FTDA only, as COALLIA data collection is still ongoing.

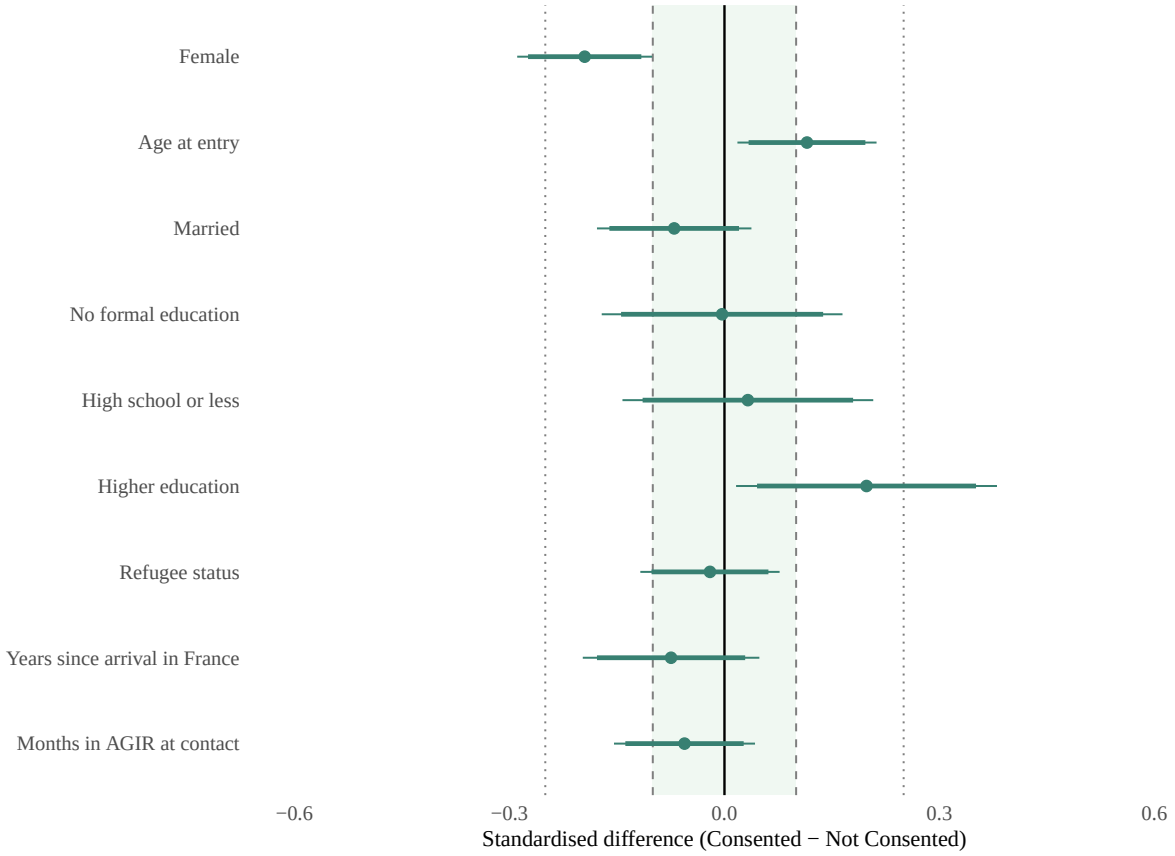

Figure A9: Covariate Balance by Consent Status Among Individuals Seen by RISE

Notes: For each baseline characteristic, the figure plots the standardised difference between individuals who consented and those who did not consent to participate in RISE, computed as the consented-minus-not-consented mean divided by the not-consented-group standard deviation. The sample is restricted to individuals seen by RISE. Individuals who requested additional time to decide ("Wanted to think") are grouped with those who explicitly did not consent. Thick bars are 90% confidence intervals; thin bars are 95% confidence intervals. The inner dashed lines and shaded band delimit the $[-0.1, 0.1]$ range, within which standardised differences are typically considered negligible. The outer dotted lines delimit the $[-0.25, 0.25]$ range. *Years since arrival in France*: difference in years between France arrival date and AGIR programme entry date. *Months in AGIR at contact*: difference in months between the date where the RISE team met the person, and the AGIR programme entry date. Observations with missing values are dropped on a per-variable basis. We note that the variables used in this figure are those extracted automatically from AGIR's information system. In the dataset used for analysis containing consenting individuals only, these variables have been corrected and supplemented, when missing, through manual data collection.

B Data and Variable Construction

The three sources are (i) monitoring data collected by the mental-health partner structures hosting RISE-oriented refugees, (ii) admissions records from the Assistance Publique - Hôpitaux de Paris (AP-HP) hospital information system (ORBIS), and (iii) the call log of the Helpline operated by partner mediators. For each refugee in the sample, each outcome is constructed by matching records from these sources to the individual's pseudo-anonymous identifier and splitting events in time around the individual's random-assignment date (the *treatment date*). Events dated on or before the treatment date feed baseline controls; events dated strictly after feed post-treatment outcomes. The three sources were last extracted on 8 July 2026. When a refugee has no record in a given source, the corresponding count is set to zero rather than missing, consistent with the administrative interpretation that an absent record reflects the absence of the event.

B.1 Outcomes

Any partner-structure appointment after treatment The partner-structure monitoring data record, for each RISE-oriented appointment, the date, the referring structure, the appointment type (psychiatric consultation, general medical consultation, or administrative), and its status (held, cancelled by the patient, missed, or phone consultation). For each refugee, the indicator *any partner-structure appointment* equals 1 if at least one appointment dated after the treatment date is recorded in any partner structure, and 0 otherwise (including refugees with no record at all). Appointments with missing dates are dropped from the count.

Number of psy appointments after treatment Among all appointments dated after the treatment date, the subset flagged as psychiatric consultations is counted for each refugee. Cancelled, missed, and phone consultations are included in this count provided they carry a date; this choice follows the partner structures' own definition of a psy appointment as any scheduled event with a psychiatrist, regardless of attendance status. The variable is zero for refugees with no post-treatment psy appointment on record.

Number of appointments after treatment (any type) For each refugee, the total number of appointments with any partner structure dated strictly after the treatment date, across all appointment types (psychiatric, medical, administrative). The variable is zero for refugees with no post-treatment record.

Any hospital visit after treatment The AP-HP hospital monitoring data record hospital-admission dates per refugee, corresponding to the first distinct admissions to a particular hospital found in ORBIS. The indicator *any hospital visit* equals 1 if at least one admission date is strictly greater than the treatment date, and 0 otherwise.

Number of hospital visits after treatment For each refugee, the count of distinct admission dates strictly greater than the treatment date. Refugees with no admission record after treatment have a count of zero.

Called the helpline The helpline monitoring data record each call received by the dedicated Helpline number, including the caller's pseudonymised identifier when it could be matched to the study sample, the date of the call, its language, whether the call was motivated by a mental-health concern, and the mediator's action. The indicator *called the helpline* equals 1 if at least one call-record row is matched to

the refugee's identifier with a non-missing call date, and 0 otherwise. By construction, refugees assigned to arms without helpline access (Control, Peer-Helper) have no exposure to the number and their value is zero.[44]

Called the Helpline for a Psychiatric Reason. Among refugees who called the helpline, this indicator equals 1 if the *reason for call* field recorded by the mediator identifies a mental-health concern, and 0 otherwise. Refugees who never called the helpline are coded as 0. Calls are classified as mental-health related if the refugee specifically requested an appointment with a psychologist or psychiatrist, as opposed to calls made for general information or to book an appointment with a non-psychiatric medical professional.

Helpline secured an appointment Among refugees who called, this indicator equals 1 if the mediator recorded that a downstream appointment with a partner mental-health structure was secured during or immediately after the call, and 0 otherwise. Refugees who did not call, or called for non-psychiatric appointment are coded as 0 by construction.

Attended the Helpline-Secured Appointment. Conditional on a helpline-secured appointment, this indicator equals 1 if the partner structure or the refugee subsequently confirmed attendance, and 0 if the refugee did not show up or no confirmation was recorded. Refugees for whom no appointment was secured are coded as 0 by construction.[45]

Early exit from the AGIR programme The AGIR programme has a mandatory maximum duration of 24 months from the refugee's enrolment date. FTDA's monitoring system records, for each refugee, the date and motive of exit whenever an exit has occurred. The indicator *early exit from AGIR* equals 1 if the refugee has an exit dated strictly after the treatment date *and* the recorded motive is anything other than *Reached maximum support period* (the motive used by the monitoring system to flag refugees whose 24-month cap has been reached). The indicator equals 0 for refugees still in the programme at the collection cut-off and for refugees who exited solely because of the cap. By FTDA's convention, cap exits are recorded at the end of the calendar month preceding the expected cap date (e.g., an expected cap on 2025-12-06 is recorded as an exit on 2025-11-30).

Positive exit from the AGIR programme Restrictive version of the preceding outcome. The indicator *positive exit* equals 1 if the refugee exited after the treatment date *and* the recorded motive is one of the five motives classified as positive by FTDA: *Access to permanent housing and stable employment*; *Access to permanent housing — Private housing*; *Access to permanent housing — Social housing*; *Access to permanent housing — Other*; or *Access to stable employment*. The indicator equals 0 for all other refugees, including those who exited for administrative, transitional, or negative reasons. Refugees still in the programme at the cut-off are coded 0 (right-censored observations cannot be positive by construction).

Employment and Housing Trajectories The manual data collection traces, for every refugee, the complete sequence of employment and housing spells from entry into the AGIR programme onwards, each spell being characterised by a start date, an end date and a category. Trajectory variables are built from these spells as calendar-day counts over two observation windows: a pre-treatment window running from AGIR entry to the day before treatment, and a post-treatment window running from the treatment

[44] No crossovers are observed in the data, i.e. no refugee assigned to the control or peer-helper arms called the helpline.

[45] Mediators call the refugee the day following their appointment to confirm attendance.

date to the earlier of the AGIR exit and the measurement date of 28 July 2026. Days are counted as the union of the relevant spells, so that overlapping spells, such as two concurrent part-time contracts, never contribute a day twice; spell end dates are inclusive. Each count is then divided by the length of its window, which yields shares bounded between zero and one by construction. A refugee whose window contains no categorised spell is missing on the corresponding variable rather than counted as zero.

Six trajectory variables are used. On the extensive margin, *active* pools every employment, education and training category, that is, all recorded situations except unemployment; *employed* covers actual work contracts only, excluding students and trainees; and *stable housing* pools private or personal accommodation, social housing and residences, as opposed to accommodation with third parties, reception centres, hotels and street or shelter situations. On the intensive margin, *stable employment* retains open-ended contracts and fixed-term contracts of more than six months, including their integration variants; *full-time employment* uses the full-time flag recorded alongside each work spell; and *own-name housing* retains private and social accommodation held in the refugee's or the spouse's name, the most autonomous of the housing arrangements.

Categories of Exit from the AGIR Programme Exits are classified from the motive recorded by the programme's caseworkers. Because the recording vocabulary was revised during the study, each family below pools the historical wording and its current equivalent. *Positive* exits gather departures for permanent housing, for stable employment, and entries into a temporary accommodation centre for recognised refugees (CPH, *centre provisoire d'hébergement*), a dedicated follow-on housing scheme that caseworkers record as the exit reason only when the housing solution was deemed sufficient. *Administrative cap* gathers the two wordings of the twenty-four-month ceiling, which the partner records at the end of the month preceding the expected cap date. *Other early exits* gather departmental moves, referrals to local actors, emergency or supported accommodation, exclusions, dropouts and residual motives.

Together with the refugees still enrolled at the measurement date, these families form a partition of the sample: every refugee falls into exactly one of *still enrolled*, *reached the cap*, *exited positively* and *exited otherwise*, so that the four intention-to-treat coefficients of a given arm sum to zero by construction. Table B1 reports the distribution of the individual motives across arms, Table B2 checks the classification against the collected trajectories at the exact exit date, and Table B3 describes the composition of exit behaviour by arm.

Table B1: Distribution of AGIR exit motives across treatment arms

Exit motif	Control	Peer-Helper	Helpline	Peer-Helper + Helpline	Total
Positive (housing, employment or CPH)					
End of support - Access to permanent housing - Other	0	1	0	0	1
End of support - Access to permanent housing - Private housing	0	0	1	0	1
End of support - Access to permanent housing and stable employment	12	8	5	6	31
Positive exit: Housing	16	8	12	8	44
Simple exit - Entry into CPH (permanent housing program)	1	1	5	1	8
Simple exit: Entry into CPH	2	3	2	2	9
Intermediate (referral to a local actor)					
Simple exit - Relocation and referral to local actor	1	4	3	2	10
Administrative cap (24-month ceiling)					
24 months of support reached	80	76	59	84	299
Reached maximum support period	33	34	27	27	121
Negative / other					
Abandonment	0	1	0	0	1
Exclusion	1	0	2	0	3
Exclusion - Refusal of adapted housing	0	0	0	1	1
Simple exit - CHRS (emergency shelter)	2	2	0	1	5
Simple exit - CHS with support (residential facility)	1	1	0	0	2
Simple exit - Other	1	1	0	0	2
Simple exit: Moving out of the department	4	3	2	6	15
Simple exit: Other	2	1	1	0	4

Notes: This table reports the number of refugees in each treatment arm recorded with each of the 14 observed AGIR exit motives, as of the 28 July 2026 collection. Rows are grouped into four categories: *Positive* = refugee left the programme having obtained stable housing and/or stable employment, or entered a temporary accommodation centre for recognised refugees (CPH); *Intermediate* = refugee was referred to a local actor; *Administrative cap* = refugee reached the 24-month ceiling of AGIR support (recorded at the end of the month preceding the expected cap date); *Negative / other* = residual motives (dropout, exclusion, emergency shelter, unspecified). *Early exit from AGIR* counts any exit outside the Administrative cap category, and *Positive exit* the Positive category only. The Intermediate category is shown for transparency but is not a reported outcome.

Table B2: Exit categories validated against the collected trajectories

Exit category	N	Median months in AGIR at exit	Stable housing at exit (%)	Employed at exit (%)	Stably employed at exit (%)
Positive exit	93	20.7	94	77	68
Reached the 24-month cap	419	23.5	15	64	48
Other early exit	48	18.6	35	25	23

Notes: Each row describes one exit category among the refugees with a post-treatment AGIR exit. Positive exits pool the housing, employment and CPH-entry motives; the cap category corresponds to the administrative 24-month ceiling; other early exits pool departmental moves, referrals, emergency accommodation, exclusions and dropouts. Housing and employment status are read off the manually collected trajectory spells covering the exact exit date: stable housing pools private, personal and social housing or residences; stable employment pools open-ended and six-month-plus contracts.

Table B3: Exit composition by arm (descriptive)

Arm	N	Still enrolled (%)	Exits	Cap among exits (%)	Early exits	Positive among early (%)
Control	290	46.2	156	72.4	43	72.1
Peer-Helper	286	49.3	145	75.2	36	58.3
Helpline	274	56.2	120	71.7	34	70.6
Peer-Helper + Helpline	285	51.2	139	79.9	28	60.7

Notes: Descriptive composition of exit behaviour by treatment arm over the observation window (treatment to AGIR exit or the measurement date). The conditional shares (cap among exits, positive among early exits) condition on exiting, which is itself affected by treatment; they describe composition and carry no causal interpretation, unlike the intention-to-treat estimates on the four unconditional categories reported below.

B.2 Control Variables

The following covariates are used as baseline controls in the specifications that include a covariate pool (Spec 2: strata FE plus the full pool; Spec 3: strata FE plus the pool's Belloni-Chernozhukov-Hansen double-selection LASSO-selected subset). For every covariate with missing values we include a separate missingness indicator and recode the missing cell to a neutral placeholder, so the coefficient on the original covariate is identified only on the observed sub-sample while no observation is dropped from the regression. Categorical variables enter as factor indicators with an explicit reference category.

Demographics and migration history

- **Age at entry**: Age in years at the refugee's administrative entry date into the AGIR programme.
- **Female**: Binary indicator equal to 1 for female refugees and 0 otherwise. Used in the strata variable (see below) rather than as a separate regressor.
- **Nationality group**: Nine-level categorical derived from declared nationality. Refugees from the four nationalities with at least 40 observations each (Afghan, Bangladeshi, Ivorian, Sri Lankan) form four standalone groups; all remaining nationalities are pooled into five regional groups following the World Bank regional classification: Sub-Saharan Africa, Middle East and North Africa, Asia (excluding the four standalone Asian nationalities), Europe and Central Asia, and Latin America and Caribbean. Used in the strata variable.
- **Strata**: Interaction of nationality group and female, yielding 18 non-empty cells. All specifications absorb strata fixed effects; strata enter neither as a regressor with reported coefficient nor as a cluster variable (standard errors are HC1 heteroskedasticity-robust throughout).
- **Years since arrival in France**: Difference between the refugee's administrative entry date into AGIR and the treatment date, expressed in years.
- **Months in AGIR at treatment**: Difference between the treatment date and the AGIR entry date, expressed in months.
- **Months since CIR signature**: Difference between the treatment date and the date at which the refugee signed the French Contrat d'Integration Republicaine with the *Office Français de l'Immigration et de l'Intégration (OFII)*, expressed in months.
- **Months since protection grant**: Difference between the treatment date and the date at which the refugee was granted international protection status by OFPRA or the CNDA, expressed in months. Available for 94.8% of the sample; a paired missingness indicator is included in regressions using this covariate.

Baseline mental health

- **PTSD symptom score**: Sum of the six International Trauma Questionnaire (ITQ) symptom items covering re-experiencing, avoidance, and sense of current threat. Range 0–24; defined only when all six constituent items are observed. See B.3 for exact item wording, response scales, and scoring rules.
- **DSO symptom score**: Sum of the six ITQ items covering affective dysregulation, negative self-concept, and disturbed relationships. Range 0–24; defined only when all six constituent items are observed. See B.3.

- **Depression-screen score**: Sum of the two depression items administered in the RISE baseline (items 1 and 2 of the International Depression Questionnaire, equivalent to the Patient Health Questionnaire-2 or PHQ-2). Range 0–6; defined only when both items are observed. See B.3.

The two summary indices are constructed from the four Likert items of the baseline self-administered questionnaire. Each index aggregates two items measured on the same direction (higher = better) but on different scales; it provides a compact one-dimensional measure of an underlying construct and allows treatment effects to be expressed in standard-deviation units of the Control arm. Both indices are built with the same three-step procedure. First, each component item is standardised on the Control-arm mean and standard deviation, so each standardised item has mean 0 and standard deviation 1 in the Control arm. Second, the two standardised items are averaged with equal weight on each respondent, using all available (non-missing) items; respondents with both items missing are coded as missing on the index. Third, the averaged score is itself re-standardised on the Control-arm distribution so that the final index has Control-arm mean 0 and standard deviation 1 by construction. Treatment effects on either index are therefore interpretable directly as a fraction of a Control-arm standard deviation.

Self-reported health index. This index combines two items on subjective health, asked in succession at the start of the health module.

1. *"How is your health in general?"* Response options: very bad, bad, average, good, very good ("Do not wish to answer" coded as missing). Coded on a five-point scale from 1 (very bad) to 5 (very good).
2. *"How satisfied are you with your current health?"* Response options: not at all satisfied, not really satisfied, quite satisfied, very satisfied ("Do not wish to answer" coded as missing). Coded on a four-point scale from 1 (not at all satisfied) to 4 (very satisfied).

Healthcare-system integration index. This index combines two items on refugees' ability to navigate and access the French healthcare system.

1. *"Would you say that you know the healthcare system in France well?"* Response options: I don't know it at all, I know it a little, I know it well, I know it very well ("Do not wish to answer" coded as missing). Coded on a four-point scale from 1 (I don't know it at all) to 4 (I know it very well).
2. *"When you need care, is it easy for you to get an appointment with a healthcare professional?"* Response options: very difficult, rather difficult, pretty easy, very easy ("Do not wish to answer" coded as missing). Coded on a four-point scale from 1 (very difficult) to 4 (very easy).

Pre-treatment healthcare utilisation

- **Any prior partner-structure contact**: Binary indicator equal to 1 if the refugee has at least one recorded appointment with a partner mental-health structure dated on or before the treatment date, and 0 otherwise.
- **Number of pre-treatment hospital visits**: Count of AP-HP admission dates dated on or before the treatment date. Refugees with no prior admission record have a count of zero.

Education Derived from the self-reported highest qualification level at AGIR entry, available for 1,036 of the 1,135 refugees in the estimation sample. Missingness is absorbed by a paired indicator in the regression.

- **No formal education**: Binary indicator equal to 1 if the declared qualification level is "no formal education" and 0 if any formal qualification is declared.
- **Has higher education**: Binary indicator equal to 1 if the declared qualification level corresponds to any post-secondary qualification (some college, professional degree, bachelor's, postgraduate diploma, master's, or doctorate), and 0 otherwise.

Housing and civic-integration-contract intensity

- **Stable housing at entry**: Binary indicator equal to 1 if the housing situation at AGIR entry corresponds to a stable arrangement (private rental or social housing, aggregated into a single category in the coded housing variable), and 0 for precarious arrangements (hotel, shelter, staying with family or friends). Coded as missing when the housing situation is unknown.
- **CIR French-language hours**: Number of French-language hours prescribed in the civic-integration contract signed with OFII. A value of zero indicates that no hours were prescribed (the refugee was assessed above the target proficiency at signature).
- **Stable housing on the treatment day**: Binary indicator equal to 1 if a housing spell classified as stable covers the treatment date, 0 if a classified spell covers it but corresponds to a precarious arrangement, and missing if no classified spell covers that date. Unlike the previous indicator, which refers to the situation at entry into the programme, this one measures the situation at randomisation, which is the relevant baseline for outcomes recorded afterwards and, mechanically, for exits that require a housing solution.
- **In employment on the treatment day**: Binary indicator built symmetrically from the employment spells, equal to 1 if a work contract covers the treatment date.

B.3 Baseline Mental-Health Instruments

This subsection documents the two validated self-report instruments used at baseline to measure mental-health symptomatology: the International Trauma Questionnaire (ITQ), which screens for post-traumatic stress disorder (PTSD) and complex PTSD following the International Classification of Diseases 11th revision (Cloitre et al., 2018); and the depression screen drawn from items 1 and 2 of the International Depression Questionnaire (IDQ), equivalent to the widely-used two-item Patient Health Questionnaire (Kroenke et al., 2003; Shevlin et al., 2023). For each instrument we reproduce the exact item wording administered to the refugees, the response scale, and the dimensional and diagnostic scoring rules pre-specified in the scale-development papers. We also flag the one item cluster that was not collected in the RISE baseline and explain its implication for scoring.

B.3.1 International Trauma Questionnaire (ITQ): PTSD symptoms

Respondents were told: *"We are now going to ask you some questions about the reactions, emotions, or feelings you may experience after having lived through a potentially traumatic event. We are not asking you to share your personal story or the difficult things you may have experienced. However, we would like to know whether, today, your past experiences are creating difficulties for you in your daily life. Would you say that you sometimes:"* Each of the six PTSD symptom items is rated on a five-point Likert scale: (0) "not at all," (1) "a little bit," (2) "moderately," (3) "quite a bit," (4) "extremely." "Do not wish to answer" responses are coded as missing.

Re-experiencing in the here and now (items P1–P2)

1. *"Have disturbing dreams related to difficult past events that you experienced?"*
2. *"Have intense images or memories that give you the feeling of reliving difficult moments from the past?"*

Avoidance (items P3–P4)

3. *"Avoid feelings that remind you of difficult moments (for example, thoughts, emotions, or physical sensations)?"*
4. *"Avoid external things that remind you of difficult moments (for example, people, places, conversations, objects, activities, or situations)?"*

Sense of current threat (items P5–P6)

5. *"Be in a constant state of alert, excessive vigilance, or always on guard?"*
6. *"Have exaggerated startle or surprise reactions?"*

B.3.2 ITQ: PTSD functional impairment

Respondents were then asked: *"Over the past month, these emotional difficulties, beliefs about yourself, or your relationships have :"* with the same five-point Likert scale.

7. *"Affected your relationships and social life?"*
8. *"Affected your work or ability to work?"*
9. *"Affected other important parts of your life such as the ability to care for your children, your studies, or any other important activities?"*

B.3.3 ITQ: Disturbances in Self-Organisation (DSO) symptoms

Respondents were told: *"Below is a list of problems or symptoms that people who have experienced a stressful event may sometimes experience. Answer how true the statement is about you:"* The five-point Likert scale here is slightly differently phrased but numerically identical: (0) "not at all," (1) "a little bit," (2) "moderately," (3) "quite a bit," (4) "extremely."

Affective dysregulation (items C1–C2)

10. *"When I am upset, it takes me a long time to calm down."*
11. *"I feel numb or emotionally shut down."*

Negative self-concept (items C3–C4)

12. *"I feel like a failure."*
13. *"I feel worthless."*

Disturbances in relationships (items C5–C6)

14. *"I feel distant or cut off from others"*
15. *"I find it difficult to stay emotionally close to others"*

B.3.4 ITQ: DSO functional impairment (not collected at baseline)

The ITQ's DSO-functional-impairment cluster (items C7–C9) was not administered in the RISE baseline questionnaire. For completeness the items are reproduced here; the RISE baseline stops at C6. Consequently the sample permits the computation of a DSO symptom score and DSO symptom-endorsement flags, but the full complex-PTSD diagnosis per the ITQ algorithm (which requires endorsement of at least one of C7–C9) *cannot* be computed from the RISE baseline data. This limitation is flagged in the descriptive-statistics tables in B.4.

C7. *"Created concern or distress about your relationships or social life?"* — not collected.

C8. *"Affected your work or ability to work?"* — not collected.

C9. *"Affected any other important parts of your life such as parenting, or school or college work, or other important activities?"* — not collected.

B.3.5 ITQ: Scoring rules

Symptom-cluster endorsement. An ITQ symptom cluster is *endorsed* when at least one of its two items scores ≥ 2 (*i.e.* "moderately," "quite a bit," or "extremely"). When both items of a cluster are missing, the endorsement flag is missing; when only one is missing but the observed one is ≥ 2, the cluster is endorsed. The six cluster-endorsement flags are computed accordingly: Re-experiencing, Avoidance and Threat for the PTSD side, and Affective dysregulation, Negative self-concept and Disturbed relationships for the DSO side.

PTSD functional-impairment flag. The PTSD functional-impairment criterion is met when at least one of P_7, P_8, P_9 is ≥ 2.

PTSD diagnosis. The refugee meets ICD-11 PTSD criteria when Re-experiencing AND Avoidance AND Threat AND the PTSD functional-impairment flag are all met.

Complex-PTSD diagnosis. The RISE baseline does not collect DSO functional-impairment items (C7–C9), so a full complex-PTSD diagnosis per the ITQ algorithm cannot be computed. As a partial substitute we report a "DSO symptom-criteria met" flag that is positive when Affective dysregulation AND Negative self-concept AND Disturbed relationships are all endorsed, tagging refugees who meet the DSO symptom component but whose functional-impairment status in the DSO domain is unknown. This flag is a necessary but not sufficient condition for complex PTSD.

B.3.6 Depression screen (IDQ items 1–2, equivalent to PHQ-2)

The RISE baseline administered the two items from the Affective cluster of the International Depression Questionnaire that coincide item-for-item with the Patient Health Questionnaire-2 (Kroenke et al., 2003; Shevlin et al., 2023). Respondents were asked: *"Over the past two weeks, how often have you been bothered by the following problems?"*. Unlike the ITQ's five-point scale, the depression items use a four-point frequency scale: (0) "never," (1) "several days," (2) "more than seven days," (3) "almost every day."

1. *"I find no interest or pleasure in doing things"*
2. *"I have a feeling of sadness, depression or hopelessness"*

Depression-screen score. The depression score is the sum of the two items, ranging from 0 to 6 and defined only when both items are observed. Higher values indicate more severe depressive symptomatology.

B.4 Baseline Mental Health

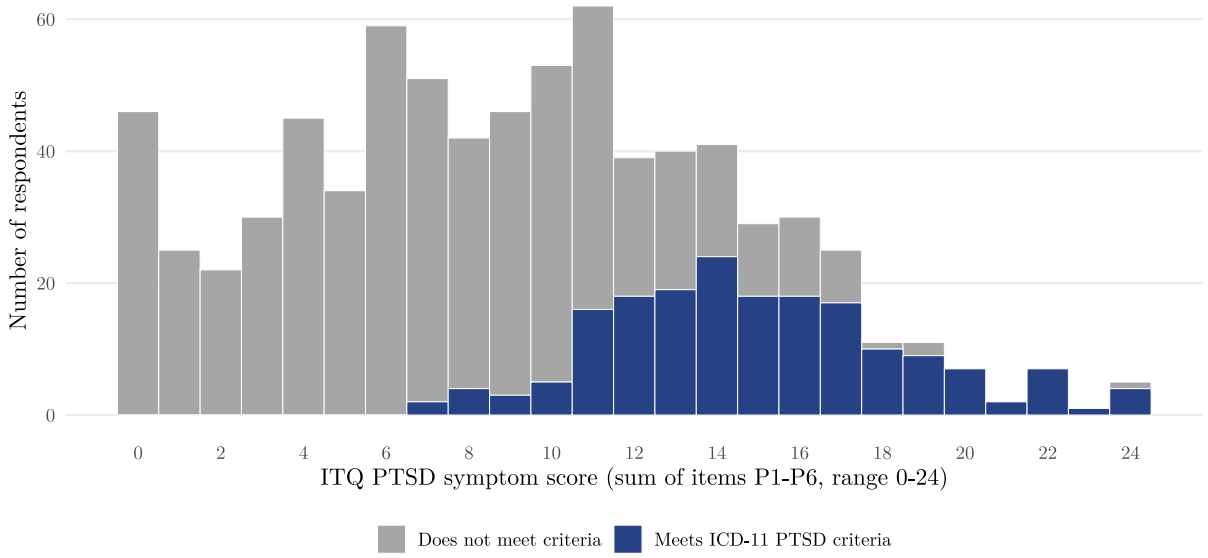

Figure B1: Distribution of ITQ PTSD Symptom Scores at Baseline

Notes: Histogram of the ITQ PTSD symptom score (sum of items P_1-P_6, range 0-24) for respondents with a non-missing score on all six symptom items. Bar colour indicates whether the respondent also meets the full ICD-11 PTSD diagnostic criteria.

Table B4: ITQ Item Distribution and Diagnostic Construction

		Likert response (or 0/1 for diagnostics)					
	N	0	1	2	3	4	DNWA
PTSD symptoms (P1-P6)							
P1: Disturbing dreams	1005	285	364	163	139	54	124
P2: Reliving images	1008	282	338	172	167	49	117
Re-experiencing (Re = 1 if P1 or P2 $\geq$ 2)	1058	550	508 (48.0%)				
P3: Internal avoidance	945	197	285	167	174	122	184
P4: External avoidance	957	223	258	161	195	120	169
Avoidance (Av = 1 if P3 or P4 $\geq$ 2)	1023	393	630 (61.6%)				
P5: Hypervigilance	972	254	215	200	196	107	157
P6: Exaggerated startle	985	374	300	146	105	60	142
Sense of threat (Th = 1 if P5 or P6 $\geq$ 2)	1044	444	600 (57.5%)				
PTSD functional impairment (P7-P9)							
P7: Affected social life	950	398	265	103	117	67	176
P8: Affected work	987	449	286	99	101	52	141
P9: Affected parenting/other	923	440	212	112	99	60	206
Functional impairment (FI = 1 if P7/P8/P9 $\geq$ 2)	1038	571	467 (45.0%)				
PTSD diagnosis							
PTSD diagnosis (= 1 if Re AND Av AND Th AND FI)	954	750	204 (21.4%)				
Disturbances in Self-Organization - Affective dysregulation (C1-C2)							
C1: Difficulty calming down	1007	237	351	155	167	97	118
C2: Emotional numbness	969	410	241	131	112	75	158
Affective dysregulation (AD = 1 if C1 or C2 $\geq$ 2)	1046	518	528 (50.5%)				
Disturbances in Self-Organization - Negative self-concept (C3-C4)							
C3: Feeling like a failure	987	599	164	75	86	63	138
C4: Feeling worthless	985	629	153	57	96	50	141
Negative self-concept (NSC = 1 if C3 or C4 $\geq$ 2)	1020	740	280 (27.5%)				
Disturbances in Self-Organization - Disturbed relationships (C5-C6)							
C5: Feeling distant	996	442	259	97	122	76	130
C6: Difficulty staying close	964	435	233	114	106	76	162
Disturbed relationships (DR = 1 if C5 or C6 $\geq$ 2)	1030	623	407 (39.5%)				

Notes: This table presents the response distribution for each ITQ item and the constructed diagnostic variables. Likert scale: 0 = Not at all, 1 = A little bit, 2 = Moderately, 3 = Quite a bit, 4 = Extremely. DNWA = Do not wish to answer (treated as missing). Diagnostic components are binary: endorsed (= 1) when at least one item in the cluster scores $\geq$ 2. If one item is missing but the other scores $\geq$ 2, the component is endorsed. If all items are missing, the component is NA and excluded from N. PTSD is diagnosed when all four components (Re, Av, Th, PTSD-FI) are endorsed. Disturbances-in-Self-Organization components (AD, NSC, DR) are reported individually. Disturbances-in-Self-Organization functional impairment items (C7-C9) were not collected in the baseline survey, so neither the Disturbances-in-Self-Organization diagnosis nor the complex-PTSD diagnosis can be computed.

Table B5: Evaluability per ITQ Diagnostic Component

Diagnostic component	% Evaluable	% Non-evaluable
Re-experiencing (RE)	93.2	6.8
Avoidance (AV)	90.1	9.9
Threat perception (TH)	92.0	8.0
Functional impairment (PTSD-FI)	91.5	8.5
Affective dysregulation (AD)	92.2	7.8
Negative self-concept (NSC)	89.9	10.1
Disturbed relationships (DR)	90.7	9.3

Notes: A diagnostic component is evaluable when at least one of its constituent items has a valid response (i.e., not missing and not "Do not wish to answer"). Percentages are computed over the full analysis sample (N = 1135).

Table B6: PTSD Prevalence by Missingness Group (Diagnosable Cases Only)

Missingness group	N respondents	N PTSD	% PTSD
All diagnosable	954	204	21.4
0 missing	675	165	24.4
1 missing	173	31	17.9
2 missing	68	5	7.4
3+ missing	38	3	7.9

Notes: An individual is diagnosable for PTSD when at least one item is observed in each of the four required clusters (re-experiencing, avoidance, sense of threat, functional impairment). PTSD is diagnosed when all four criteria are met (score ≥ 2 on at least one item per cluster). Missingness refers to the number of missing items among the nine PTSD items (P1-P9).

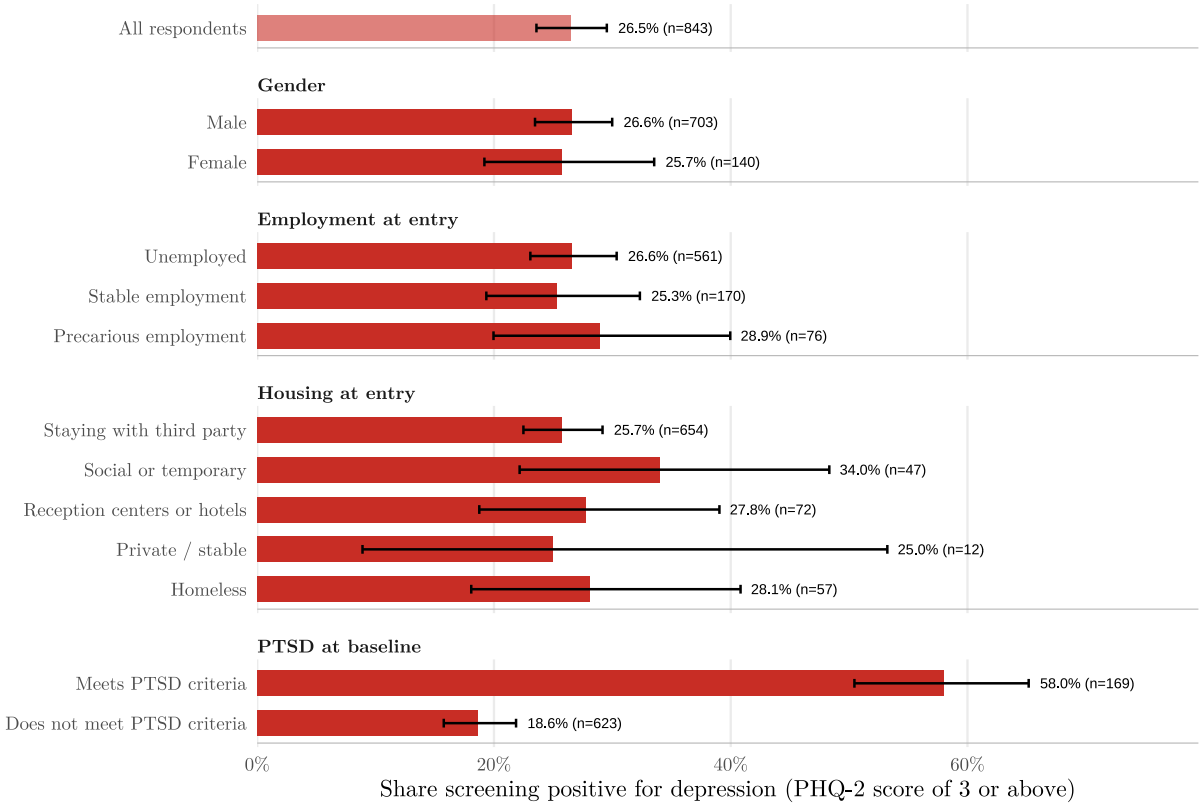

Figure B2: Baseline depression-screen positive rate, by subgroup

Notes: Each bar reports the share of baseline respondents in the indicated subgroup with a PHQ-2 score of 3 or above (sum of two items on feeling down and loss of interest, each scored 0-3). Horizontal whiskers are Wilson 95% confidence intervals around the subgroup share and are therefore not centred on the observed proportion.

C Balance and Baseline Characteristics

Table C1: Balance tests on baseline characteristics across randomised groups

	Mean of variable by treatment				P-value for the test of equality of means						Signif. under MHT
	Control	Peer-Helper	Helpline	PH+Helpline	PH vs C	HL vs C	PH+HL vs C	PH vs HL	PH+HL vs HL	PH+HL vs PH	
Demographics											
Female	0.20	0.17	0.18	0.15	0.50	0.67	0.12	0.81	0.26	0.37	No
Age at entry	31.30	32.35	30.23	30.93	0.23	0.05*	0.52	0.00***	0.17	0.05*	No
Married	0.27	0.21	0.21	0.20	0.08*	0.09*	0.05*	0.97	0.82	0.85	No
Number of children	1.06	1.04	0.97	0.92	0.90	0.53	0.29	0.46	0.65	0.24	No
Head of household	0.94	0.95	0.95	0.96	0.78	0.77	0.92	0.97	0.81	0.83	No
No formal education	0.22	0.14	0.15	0.14	0.04**	0.05*	0.03**	0.96	0.82	0.86	No
High school or less	0.55	0.59	0.68	0.59	0.25	0.00***	0.38	0.08*	0.04**	0.79	No
Higher education	0.23	0.27	0.17	0.27	0.71	0.05*	0.40	0.03**	0.01***	0.65	No
French language hours (CIR)	283.75	263.93	266.92	264.29	0.50	0.54	0.43	0.97	0.87	0.89	No
Migration background											
Refugee status	0.92	0.93	0.92	0.94	0.45	0.91	0.31	0.39	0.26	0.81	No
Years since arrival in France	3.20	3.07	3.21	3.35	0.42	0.99	0.39	0.40	0.39	0.12	No
Months in AGIR at treatment	11.91	11.96	11.41	11.89	0.82	0.27	0.91	0.18	0.32	0.74	No
Months since CIR signature	15.48	15.86	14.74	15.56	0.38	0.23	0.83	0.03**	0.15	0.50	No
Employment at entry											
Stable	0.20	0.20	0.19	0.19	0.79	0.91	0.99	0.71	0.90	0.80	No
Precarious	0.08	0.10	0.10	0.10	0.32	0.38	0.32	0.92	0.91	0.99	No
Unemployed	0.73	0.70	0.71	0.71	0.39	0.63	0.51	0.70	0.85	0.84	No
Active (employed, student or training)	0.30	0.32	0.31	0.32	0.56	0.74	0.62	0.81	0.87	0.93	No
Housing at entry											
Private / stable	0.01	0.02	0.01	0.01	0.59	0.65	0.87	0.32	0.78	0.48	No
Staying with third party	0.72	0.80	0.78	0.81	0.04**	0.16	0.01**	0.55	0.29	0.63	No
Reception centers or hotels	0.11	0.09	0.07	0.07	0.81	0.27	0.20	0.37	0.85	0.29	No
Social or temporary	0.06	0.04	0.05	0.05	0.19	0.37	0.59	0.70	0.69	0.42	No
Homeless	0.09	0.05	0.09	0.06	0.05**	0.93	0.05*	0.06*	0.07*	0.96	No
Baseline trajectories											
Share of days active	0.49	0.49	0.48	0.50	0.68	0.86	0.71	0.57	0.58	0.96	No
Share of days employed	0.40	0.44	0.41	0.42	0.17	0.60	0.50	0.40	0.89	0.47	No
Share of days in stable housing	0.08	0.09	0.05	0.07	0.99	0.10	0.60	0.13	0.30	0.63	No
Situation on the treatment day											
In stable housing	0.14	0.15	0.11	0.11	0.98	0.40	0.24	0.41	0.74	0.25	No
In employment	0.54	0.52	0.52	0.52	0.72	0.55	0.62	0.82	0.93	0.89	No
Self-reported health-related indices											
Self-reported health index	0.00	-0.07	-0.06	-0.03	0.49	0.40	0.51	0.89	0.87	0.98	No
Healthcare-system integration index	0.00	-0.13	-0.21	-0.11	0.05**	0.01***	0.09*	0.46	0.31	0.77	No

Table C1: Balance tests on baseline characteristics across randomised groups *(continued)*

	Control	Peer-Helper	Helpline	PH+Helpline	PH vs C	HL vs C	PH+HL vs C	PH vs HL	PH+HL vs HL	PH+HL vs PH	
Healthcare utilisation											
Any prior contact with partner structure	0.03	0.05	0.01	0.03	0.28	0.05**	0.73	0.00***	0.03**	0.45	No
Number of psy appointments (partner structures)	0.32	0.20	0.09	0.01	0.59	0.30	0.16	0.42	0.27	0.16	No
Any hospital visit (AP-HP)	0.40	0.38	0.42	0.44	0.61	0.49	0.32	0.23	0.75	0.13	No
Number of hospital visits (AP-HP)	0.76	0.67	0.78	0.82	0.31	0.67	0.44	0.14	0.74	0.07*	No
PTSD											
Re-experiencing in the here and now	0.51	0.46	0.51	0.44	0.15	0.91	0.11	0.19	0.14	0.86	No
Avoidance	0.65	0.59	0.61	0.62	0.15	0.38	0.57	0.57	0.75	0.37	No
Sense of current threat	0.57	0.57	0.59	0.57	0.84	0.64	0.76	0.50	0.87	0.62	No
Functional impairment	0.46	0.43	0.48	0.43	0.46	0.64	0.57	0.23	0.30	0.88	No
PTSD diagnosis	0.24	0.21	0.20	0.20	0.27	0.31	0.33	0.92	0.98	0.90	No
Disturbances in Self-Organization											
Affective dysregulation	0.52	0.49	0.50	0.50	0.41	0.67	0.75	0.70	0.91	0.62	No
Negative self-concept	0.29	0.28	0.28	0.25	0.47	0.67	0.39	0.76	0.65	0.89	No
Disturbed relationships	0.44	0.36	0.38	0.40	0.06*	0.23	0.43	0.52	0.67	0.27	No
Depression											
Depression screen positive	0.29	0.22	0.27	0.28	0.05**	0.52	0.62	0.18	0.88	0.13	No

Notes: The first four columns present mean values of the baseline variable in each randomized group (Peer-Helper, Helpline, PH+Helpline = Peer-Helper + Helpline). The next six columns report p-values for pairwise tests of equality of means, estimated from a pooled regression of the variable on the three arm dummies with randomization-strata fixed effects (interaction of nationality group and gender) and HC1 heteroskedasticity-robust standard errors; arm-vs-Control p-values are read off the coefficient on each arm dummy, and arm-vs-arm p-values come from Wald tests on the contrast $\beta_i - \beta_j$ using the regression vcov. The last column reports whether the variable is still flagged as imbalanced once the p-values are corrected for multiple hypothesis testing using the Romano-Wolf stepdown procedure applied jointly to all six pairwise tests across all baseline variables in the table. The null distribution of the test statistics is simulated by permuting the treatment-group labels within randomization strata ($B = 1000$ draws), which preserves the stratified randomization design and yields an exact randomization-inference null. "Yes" flags report the strictest threshold among 10, 5 and 1 percent that rejects the null of balance for at least one of the six pairwise comparisons after Romano-Wolf adjustment. The Self-reported health and Healthcare-system integration indices are summary indices: each component is first standardised on the Control arm, the two components are averaged, and the resulting average is re-standardised on the Control arm so that each index has Control-arm mean 0 and SD 1. Stars on the raw p-values: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table C2: Descriptive Statistics by Treatment Group: Socioeconomic Characteristics at Baseline

	Full sample		Control				Peer-Helper				Helpline				PH + Help.			
	N	Mean	N	Mean	Min	Max	N	Mean	Min	Max	N	Mean	Min	Max	N	Mean	Min	Max
Demographics																		
Female	1,135	0.175	290	0.197	0.000	1.000	286	0.175	0.000	1.000	274	0.182	0.000	1.000	285	0.147	0.000	1.000
		(0.380)		(0.398)				(0.380)				(0.387)				(0.355)		
Age at entry	1,135	31.213	290	31.300	18.000	76.000	286	32.353	18.000	73.000	274	30.226	18.000	60.000	285	30.930	18.000	63.000
		(8.235)		(8.981)				(8.342)				(7.386)				(8.005)		
Married	1,119	0.223	286	0.273	0.000	1.000	284	0.211	0.000	1.000	267	0.206	0.000	1.000	282	0.202	0.000	1.000
		(0.417)		(0.446)				(0.409)				(0.405)				(0.402)		
Number of children	1,130	0.997	288	1.062	0.000	8.000	285	1.042	0.000	8.000	273	0.967	0.000	7.000	284	0.915	0.000	9.000
		(1.573)		(1.635)				(1.578)				(1.484)				(1.593)		
Head of household	1,131	0.950	287	0.941	0.000	1.000	286	0.951	0.000	1.000	274	0.953	0.000	1.000	284	0.958	0.000	1.000
		(0.217)		(0.236)				(0.216)				(0.213)				(0.202)		
No formal education	1,036	0.164	265	0.219	0.000	1.000	260	0.138	0.000	1.000	253	0.154	0.000	1.000	258	0.143	0.000	1.000
		(0.371)		(0.414)				(0.346)				(0.362)				(0.351)		
High school or less	1,036	0.601	265	0.551	0.000	1.000	260	0.592	0.000	1.000	253	0.676	0.000	1.000	258	0.589	0.000	1.000
		(0.490)		(0.498)				(0.492)				(0.469)				(0.493)		
Higher education	1,036	0.235	265	0.230	0.000	1.000	260	0.269	0.000	1.000	253	0.170	0.000	1.000	258	0.267	0.000	1.000
		(0.424)		(0.422)				(0.444)				(0.376)				(0.443)		
French language hours (CIR)	1,109	269.793	283	283.746	0.000	600.000	280	263.929	0.000	600.000	266	266.917	0.000	600.000	280	264.286	0.000	600.000
		(210.719)		(215.575)				(213.904)				(204.006)				(209.307)		
Migration background																		
Refugee status	1,135	0.927	290	0.917	0.000	1.000	286	0.934	0.000	1.000	274	0.916	0.000	1.000	285	0.940	0.000	1.000
		(0.260)		(0.276)				(0.249)				(0.278)				(0.237)		
Years since arrival in France	1,075	3.205	271	3.196	0.266	11.784	270	3.068	0.356	19.770	263	3.209	0.649	13.054	271	3.346	0.307	14.902
		(1.681)		(1.635)				(1.569)				(1.552)				(1.932)		
Months in AGIR at treatment	1,135	11.797	290	11.913	0.000	22.799	286	11.960	0.000	22.996	274	11.408	0.000	22.503	285	11.888	0.000	22.503
		(5.672)		(5.754)				(5.770)				(5.431)				(5.730)		
Months since CIR signature	1,090	15.417	275	15.482	1.347	56.767	272	15.860	1.248	33.870	264	14.739	-2.070	33.837	279	15.564	-2.070	35.217
		(6.965)		(7.177)				(7.137)				(6.615)				(6.895)		
Employment at entry																		
Stable	1,086	0.195	277	0.199	0.000	1.000	275	0.200	0.000	1.000	263	0.190	0.000	1.000	271	0.192	0.000	1.000
		(0.397)		(0.400)				(0.401)				(0.393)				(0.395)		
Precarious	1,086	0.092	277	0.076	0.000	1.000	275	0.098	0.000	1.000	263	0.095	0.000	1.000	271	0.100	0.000	1.000
		(0.289)		(0.265)				(0.298)				(0.294)				(0.300)		
Unemployed	1,086	0.713	277	0.726	0.000	1.000	275	0.702	0.000	1.000	263	0.715	0.000	1.000	271	0.708	0.000	1.000
		(0.453)		(0.447)				(0.458)				(0.452)				(0.455)		
Active (employed, student or training)	1,129	0.314	289	0.304	0.000	1.000	284	0.320	0.000	1.000	274	0.314	0.000	1.000	282	0.319	0.000	1.000
		(0.464)		(0.461)				(0.467)				(0.465)				(0.467)		
Housing at entry																		
Private / stable	1,132	0.013	290	0.014	0.000	1.000	284	0.021	0.000	1.000	274	0.007	0.000	1.000	284	0.011	0.000	1.000
		(0.114)		(0.117)				(0.144)				(0.085)				(0.102)		
Staying with third party	1,132	0.777	290	0.724	0.000	1.000	284	0.796	0.000	1.000	274	0.777	0.000	1.000	284	0.813	0.000	1.000
		(0.416)		(0.448)				(0.404)				(0.417)				(0.390)		
Reception centers or hotels	1,132	0.085	290	0.107	0.000	1.000	284	0.092	0.000	1.000	274	0.073	0.000	1.000	284	0.067	0.000	1.000
		(0.279)		(0.310)				(0.289)				(0.261)				(0.250)		
Social or temporary	1,132	0.050	290	0.062	0.000	1.000	284	0.039	0.000	1.000	274	0.047	0.000	1.000	284	0.053	0.000	1.000
		(0.219)		(0.242)				(0.193)				(0.213)				(0.224)		
Homeless	1,132	0.074	290	0.093	0.000	1.000	284	0.053	0.000	1.000	274	0.095	0.000	1.000	284	0.056	0.000	1.000

Table C2: Descriptive Statistics by Treatment Group: Socioeconomic Characteristics at Baseline *(continued)*

	N	Mean	N	Mean	Min	Max	N	Mean	Min	Max	N	Mean	Min	Max	N	Mean	Min	Max
		(0.262)		(0.291)				(0.224)				(0.294)				(0.231)		
Baseline trajectories																		
Share of days active	1,125	0.490	287	0.488	0.000	1.000	284	0.494	0.000	1.000	273	0.480	0.000	1.000	281	0.498	0.000	1.000
		(0.388)		(0.386)				(0.399)				(0.394)				(0.375)		
Share of days employed	1,125	0.419	287	0.401	0.000	1.000	284	0.440	0.000	1.000	273	0.413	0.000	1.000	281	0.420	0.000	1.000
		(0.404)		(0.401)				(0.414)				(0.407)				(0.394)		
Share of days in stable housing	1,125	0.071	287	0.080	0.000	1.000	284	0.086	0.000	1.000	273	0.050	0.000	1.000	281	0.068	0.000	1.000
		(0.227)		(0.235)				(0.252)				(0.182)				(0.232)		
Situation on the treatment day																		
In stable housing	1,129	0.128	290	0.141	0.000	1.000	282	0.149	0.000	1.000	274	0.113	0.000	1.000	283	0.106	0.000	1.000
		(0.334)		(0.349)				(0.357)				(0.317)				(0.308)		
In employment	1,128	0.526	287	0.540	0.000	1.000	286	0.521	0.000	1.000	273	0.516	0.000	1.000	282	0.525	0.000	1.000
		(0.500)		(0.499)				(0.500)				(0.501)				(0.500)		

Notes: This table reports means and standard deviations (in parentheses) of socioeconomic baseline characteristics. Binary variables are expressed as shares (multiply by 100 for percentage points). Years since arrival in France, months in AGIR and months since CIR signature are all expressed at the treatment date. French-language hours (CIR) is the number of hours prescribed in the civic-integration contract. Education dummies are built from the declared highest qualification level. Employment categories: *Stable* pools permanent contracts (CDI), fixed-term contracts longer than 6 months (CDD/CDDI $>$ 6mo) and self-employment; *Precarious* pools short fixed-term contracts (CDD/CDDI $<$ 6mo) and temping (interim); *Unemployed* is the residual. The 43 refugees recorded as students or in training at entry are left unclassified on those three dummies, which is why their sample is smaller; they are counted in *Active*. Housing categories are reported separately for each of the five recorded modalities (private or stable, staying with third party, reception centres or hotels, social or temporary, homeless). The last block reports the situation on the treatment day itself rather than at entry, read off the same spells; it is missing for the refugees with no classified spell covering that date.

Table C3: Descriptive Statistics by Treatment Group: Health and Mental-Health Status at Baseline

	Full sample		Control				Peer-Helper				Helpline				PH + Help.			
	N	Mean	N	Mean	Min	Max	N	Mean	Min	Max	N	Mean	Min	Max	N	Mean	Min	Max
Healthcare utilisation																		
Any prior contact with partner structure	1,135	0.029	290	0.028	0.000	1.000	286	0.049	0.000	1.000	274	0.007	0.000	1.000	285	0.032	0.000	1.000
		(0.168)		(0.164)				(0.216)				(0.085)				(0.175)		
Number of psy appointments (partner structures)	1,135	0.157	290	0.324	0.000	67.000	286	0.203	0.000	33.000	274	0.088	0.000	15.000	285	0.007	0.000	1.000
		(2.390)		(4.055)				(2.220)				(1.055)				(0.084)		
Any hospital visit (AP-HP)	1,135	0.408	290	0.397	0.000	1.000	286	0.378	0.000	1.000	274	0.423	0.000	1.000	285	0.435	0.000	1.000
		(0.492)		(0.490)				(0.486)				(0.495)				(0.497)		
Number of hospital visits (AP-HP)	1,135	0.759	290	0.762	0.000	7.000	286	0.675	0.000	6.000	274	0.785	0.000	6.000	285	0.818	0.000	7.000
		(1.205)		(1.229)				(1.131)				(1.214)				(1.245)		
PTSD																		
Re-experiencing in the here and now	1,058	0.480	270	0.515	0.000	1.000	265	0.460	0.000	1.000	258	0.508	0.000	1.000	265	0.438	0.000	1.000
		(0.500)		(0.501)				(0.499)				(0.501)				(0.497)		
Avoidance	1,023	0.616	255	0.647	0.000	1.000	252	0.591	0.000	1.000	255	0.608	0.000	1.000	261	0.617	0.000	1.000
		(0.487)		(0.479)				(0.493)				(0.489)				(0.487)		
Sense of current threat	1,044	0.575	268	0.575	0.000	1.000	263	0.567	0.000	1.000	256	0.590	0.000	1.000	257	0.568	0.000	1.000
		(0.495)		(0.495)				(0.496)				(0.493)				(0.496)		
Functional impairment	1,038	0.450	261	0.460	0.000	1.000	263	0.433	0.000	1.000	258	0.477	0.000	1.000	256	0.430	0.000	1.000
		(0.498)		(0.499)				(0.496)				(0.500)				(0.496)		
PTSD diagnosis	954	0.214	240	0.242	0.000	1.000	240	0.212	0.000	1.000	238	0.202	0.000	1.000	236	0.199	0.000	1.000
		(0.410)		(0.429)				(0.410)				(0.402)				(0.400)		
Disturbances in Self-Organization																		
Affective dysregulation	1,046	0.505	265	0.525	0.000	1.000	264	0.489	0.000	1.000	257	0.502	0.000	1.000	260	0.504	0.000	1.000
		(0.500)		(0.500)				(0.501)				(0.501)				(0.501)		
Negative self-concept	1,020	0.275	253	0.292	0.000	1.000	261	0.276	0.000	1.000	254	0.276	0.000	1.000	252	0.254	0.000	1.000
		(0.446)		(0.456)				(0.448)				(0.448)				(0.436)		
Disturbed relationships	1,030	0.395	254	0.437	0.000	1.000	261	0.364	0.000	1.000	253	0.383	0.000	1.000	262	0.397	0.000	1.000
		(0.489)		(0.497)				(0.482)				(0.487)				(0.490)		
Depression																		
Depression screen positive	843	0.265	209	0.292	0.000	1.000	213	0.221	0.000	1.000	207	0.266	0.000	1.000	214	0.280	0.000	1.000
		(0.441)		(0.456)				(0.416)				(0.443)				(0.450)		

Notes: This table reports means and standard deviations (in parentheses) of health- and mental-health-related baseline characteristics. General health is scored 1 (Very bad) to 5 (Very good). Health satisfaction, Knowledge of French healthcare and Ease of scheduling appointments are scored 1 to 4 (higher is better). The Self-reported health index is a summary index of General health and Health satisfaction; the Healthcare-system integration index is a summary index of Knowledge of French healthcare and Ease of scheduling. Each component is first standardised on the Control-arm mean and standard deviation, then the two components are averaged with equal weight (using all non-missing items per respondent), and the resulting average is re-standardised on the Control arm so that each index has mean 0 and standard deviation 1 in the Control arm by construction. ITQ diagnostic components (Re, Av, Th, PTSD-FI, AD, NSC, DR) are binary: endorsed when at least one item in the cluster scores ≥ 2 (Moderately or above). PTSD diagnosis requires all four PTSD components endorsed. Disturbances-in-Self-Organization functional-impairment items (C7-C9) were not collected, so neither the Disturbances-in-Self-Organization diagnosis nor the complex-PTSD diagnosis can be computed; AD, NSC and DR are reported as individual components. Depression screen positive: PHQ-2 score (sum of two items, range 0-6) of 3 or above. Healthcare utilisation variables are coded 0 for individuals not found in the partner-structure or hospital records (absence = no contact). Hospital visits come from the AP-HP information system and cover all causes. "Do not wish to answer" responses are treated as missing for survey items.

D Main Estimates by Outcome Family

D.1 Healthcare utilisation

Table D1: ITT Effects on any-Signal Outcomes

	Any post-treatment signal		
	(No cov.)	(With cov.)	(BCH-DL)
Peer-Helper	-0.010	-0.009	-0.009
	(0.020)	(0.020)	(0.020)
	[0.635]	[0.643]	[0.645]
Helpline	0.032	0.035	0.036
	(0.023)	(0.023)	(0.022)
	[0.238]	[0.182]	[0.158]
Peer-Helper + Helpline	0.075***	0.072***	0.074***
	(0.024)	(0.024)	(0.023)
	[0.007]	[0.008]	[0.005]
Covariates	No	Yes	BCH-DL
Control mean	0.062	0.062	0.062
Observations	1,135	1,135	1,135
Adjusted R^2	0.025	0.072	0.075
p: PH = Helpline	0.067	0.051	0.044
p: PH = PH + HL	0.001	0.001	0.000
p: Helpline = PH + HL	0.107	0.161	0.148
p: all jointly zero	0.003	0.003	0.001

Notes: Each column is a pooled three-arm OLS regression of the outcome on the three arm indicators with randomisation strata fixed effects (site x nationality group x gender). The three sub-columns per outcome are nested specifications: no covariates; the baseline covariate pool of the appendix on variable construction; and the subset of that pool selected by double-selection LASSO. Missing covariate values are recoded to a neutral placeholder with a paired missingness indicator, so no observation is dropped. Per arm: coefficient, HC1 robust standard error in parentheses, Benjamini-Hochberg q-value in square brackets computed within the specification. Footer: Control-arm mean, observations, pairwise tests of equality between arms, and the joint test that the three coefficients are zero. * p<0.10, ** p<0.05, *** p<0.01. BCH-DL retained the following covariates (per outcome): Any post-treatment signal: age, years in France, months in AGIR, months since CIR, months since CIR (miss.), months since protection (miss.), PHQ-2, prior contact, hosp. visits BT, higher educ., stable hous. (miss.), stable hous. at T (miss.), employed at T, employed at T (miss.).

Table D2: ITT Effects on Hospital Outcomes

	Any hospital visit (AP-HP)			Number of hospital visits (AP-HP)		
	(No cov.)	(With cov.)	(BCH-DL)	(No cov.)	(With cov.)	(BCH-DL)
Peer-Helper	-0.010	-0.010	-0.008	-0.004	-0.003	-0.001
	(0.019)	(0.019)	(0.019)	(0.023)	(0.023)	(0.023)
	[0.807]	[0.909]	[0.901]	[0.844]	[0.909]	[0.957]
Helpline	0.008	0.004	0.007	0.009	0.005	0.007
	(0.020)	(0.020)	(0.020)	(0.022)	(0.022)	(0.022)
	[0.807]	[0.909]	[0.901]	[0.807]	[0.909]	[0.901]
Peer-Helper + Helpline	0.045**	0.042*	0.045**	0.052**	0.049*	0.052**
	(0.022)	(0.022)	(0.022)	(0.025)	(0.025)	(0.025)
	[0.118]	[0.174]	[0.108]	[0.118]	[0.174]	[0.108]
Covariates	No	Yes	BCH-DL	No	Yes	BCH-DL
Control mean	0.055	0.055	0.055	0.059	0.059	0.059
Observations	1,135	1,135	1,135	1,135	1,135	1,135
Adjusted R^2	0.018	0.019	0.024	0.014	0.012	0.018
p: PH = Helpline	0.343	0.483	0.470	0.563	0.744	0.734
p: PH = PH + HL	0.012	0.019	0.015	0.036	0.061	0.047
p: Helpline = PH + HL	0.102	0.099	0.087	0.089	0.095	0.080
p: all jointly zero	0.082	0.120	0.088	0.141	0.203	0.144

Notes: Each column is a pooled three-arm OLS regression of the outcome on the three arm indicators with randomisation strata fixed effects (site x nationality group x gender). The three sub-columns per outcome are nested specifications: no covariates; the baseline covariate pool of the appendix on variable construction; and the subset of that pool selected by double-selection LASSO. Missing covariate values are recoded to a neutral placeholder with a paired missingness indicator, so no observation is dropped. Per arm: coefficient, HC1 robust standard error in parentheses, Benjamini-Hochberg q-value in square brackets computed within the specification. Footer: Control-arm mean, observations, pairwise tests of equality between arms, and the joint test that the three coefficients are zero. * p<0.10, ** p<0.05, *** p<0.01. BCH-DL retained the following covariates (per outcome): Any hospital visit (AP-HP): age, years in France, months since CIR, months since CIR (miss.), months since protection (miss.), PHQ-2, prior contact, hosp. visits BT, higher educ., stable hous. (miss.), stable hous. at T (miss.), employed at T, employed at T (miss.); Number of hospital visits (AP-HP): age, years in France, months since CIR, months since CIR (miss.), months since protection (miss.), PHQ-2, prior contact, hosp. visits BT, higher educ., stable hous. (miss.), stable hous. at T (miss.), employed at T, employed at T (miss.).

Table D3: ITT Effects on Partner-Structure Outcomes

	Any partner-structure appointment			Number of psy appointments (partner structures)			Number of appointments (partner structures, any type)		
	(No cov.)	(With cov.)	(BCH-DL)	(No cov.)	(With cov.)	(BCH-DL)	(No cov.)	(With cov.)	(BCH-DL)
Peer-Helper	0.002	-0.001	-0.001	-0.025	-0.007	-0.013	-0.021	-0.006	-0.012
	(0.008)	(0.007)	(0.007)	(0.047)	(0.027)	(0.030)	(0.047)	(0.028)	(0.030)
	[0.763]	[0.889]	[0.886]	[0.730]	[0.889]	[0.761]	[0.730]	[0.889]	[0.761]
Helpline	-0.005	0.005	0.004	-0.041	-0.009	-0.013	-0.045	-0.008	-0.013
	(0.006)	(0.005)	(0.005)	(0.044)	(0.022)	(0.023)	(0.044)	(0.022)	(0.023)
	[0.730]	[0.889]	[0.761]	[0.730]	[0.889]	[0.761]	[0.730]	[0.889]	[0.761]
Peer-Helper + Helpline	0.013	0.014*	0.013*	-0.029	-0.022	-0.027	-0.022	-0.014	-0.021
	(0.009)	(0.007)	(0.007)	(0.043)	(0.033)	(0.037)	(0.044)	(0.033)	(0.037)
	[0.730]	[0.458]	[0.622]	[0.730]	[0.889]	[0.761]	[0.730]	[0.889]	[0.761]
Covariates	No	Yes	BCH-DL	No	Yes	BCH-DL	No	Yes	BCH-DL
Control mean	0.007	0.007	0.007	0.041	0.041	0.041	0.045	0.045	0.045
Observations	1,135	1,135	1,135	1,135	1,135	1,135	1,135	1,135	1,135
Adjusted R^2	0.005	0.371	0.372	0.015	0.201	0.193	0.013	0.233	0.229
p: PH = Helpline	0.307	0.318	0.409	0.388	0.877	0.995	0.227	0.877	0.950
p: PH = PH + HL	0.279	0.083	0.097	0.825	0.537	0.544	0.965	0.728	0.726
p: Helpline = PH + HL	0.051	0.161	0.147	0.277	0.522	0.493	0.091	0.755	0.723
p: all jointly zero	0.243	0.206	0.275	0.544	0.917	0.902	0.231	0.975	0.946

Notes: Each column is a pooled three-arm OLS regression of the outcome on the three arm indicators with randomisation strata fixed effects (site x nationality group x gender). The three sub-columns per outcome are nested specifications: no covariates; the baseline covariate pool of the appendix on variable construction; and the subset of that pool selected by double-selection LASSO. Missing covariate values are recoded to a neutral placeholder with a paired missingness indicator, so no observation is dropped. Per arm: coefficient, HC1 robust standard error in parentheses, Benjamini-Hochberg q-value in square brackets computed within the specification. Footer: Control-arm mean, observations, pairwise tests of equality between arms, and the joint test that the three coefficients are zero. * p<0.10, ** p<0.05, *** p<0.01. BCH-DL retained the following covariates (per outcome): Any partner-structure appointment: age, years in France, months since CIR, months since CIR (miss.), months since protection (miss.), PHQ-2, prior contact, higher educ., stable hous. (miss.), stable hous. at T (miss.), employed at T (miss.); Number of psy appointments (partner structures): age, years in France, months since CIR, months since CIR (miss.), months since protection (miss.), PHQ-2, prior contact, higher educ., stable hous. (miss.), stable hous. at T (miss.), employed at T (miss.); Number of appointments (partner structures, any type): age, years in France, months since CIR, months since CIR (miss.), months since protection (miss.), PHQ-2, prior contact, higher educ., stable hous. (miss.), stable hous. at T (miss.), employed at T (miss.).

Table D4: ITT Effects on Helpline Outcomes

	Called the helpline			Called for a psy reason			Helpline secured an appointment			Attended the helpline-secured appointment		
	(No cov.)	(With cov.)	(BCH-DL)	(No cov.)	(With cov.)	(BCH-DL)	(No cov.)	(With cov.)	(BCH-DL)	(No cov.)	(With cov.)	(BCH-DL)
Peer-Helper	-0.003	-0.002	-0.003	-0.003	-0.002	-0.005	-0.002	-0.004	-0.004	-0.001	-0.003	-0.003
	(0.002)	(0.004)	(0.004)	(0.002)	(0.004)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.002)
	[0.372]	[0.682]	[0.441]	[0.357]	[0.543]	[0.226]	[0.388]	[0.348]	[0.226]	[0.528]	[0.373]	[0.289]
Helpline	0.030***	0.031***	0.031***	0.016**	0.018**	0.016**	0.006	0.007	0.007	0.004	0.004	0.004
	(0.011)	(0.011)	(0.011)	(0.008)	(0.008)	(0.008)	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)
	[0.017]	[0.013]	[0.014]	[0.086]	[0.070]	[0.081]	[0.302]	[0.270]	[0.226]	[0.372]	[0.373]	[0.289]
Peer-Helper + Helpline	0.043***	0.043***	0.042***	0.036***	0.037***	0.035***	0.020**	0.020***	0.019**	0.017**	0.016**	0.016**
	(0.012)	(0.012)	(0.011)	(0.011)	(0.011)	(0.011)	(0.008)	(0.008)	(0.008)	(0.008)	(0.007)	(0.007)
	[0.004]	[0.003]	[0.003]	[0.006]	[0.004]	[0.005]	[0.048]	[0.029]	[0.034]	[0.071]	[0.066]	[0.062]
Covariates	No	Yes	BCH-DL	No	Yes	BCH-DL	No	Yes	BCH-DL	No	Yes	BCH-DL
Control mean	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Observations	1,135	1,135	1,135	1,135	1,135	1,135	1,135	1,135	1,135	1,135	1,135	1,135
Adjusted R^2	0.028	0.064	0.065	0.038	0.060	0.061	0.016	0.076	0.071	0.008	0.045	0.042
p: PH = Helpline	0.003	0.002	0.002	0.033	0.024	0.024	0.180	0.084	0.090	0.343	0.198	0.190
p: PH = PH + HL	0.000	0.000	0.000	0.000	0.000	0.000	0.009	0.006	0.006	0.020	0.019	0.016
p: Helpline = PH + HL	0.454	0.450	0.461	0.145	0.153	0.153	0.157	0.159	0.167	0.122	0.145	0.147
p: all jointly zero	0.000	0.000	0.000	0.001	0.001	0.001	0.045	0.032	0.033	0.108	0.098	0.094

Notes: Each column is a pooled three-arm OLS regression of the outcome on the three arm indicators with randomisation strata fixed effects (site x nationality group x gender). The three sub-columns per outcome are nested specifications: no covariates; the baseline covariate pool of the appendix on variable construction; and the subset of that pool selected by double-selection LASSO. Missing covariate values are recoded to a neutral placeholder with a paired missingness indicator, so no observation is dropped. Per arm: coefficient, HC1 robust standard error in parentheses, Benjamini-Hochberg q-value in square brackets computed within the specification. Footer: Control-arm mean, observations, pairwise tests of equality between arms, and the joint test that the three coefficients are zero. * p<0.10, ** p<0.05, *** p<0.01. BCH-DL retained the following covariates (per outcome): Called the helpline: age, years in France, years in France (miss.), months since CIR, months since CIR (miss.), months since protection (miss.), PHQ-2, prior contact, hosp. visits BT, no educ. (miss.), higher educ., higher educ. (miss.), stable hous. (miss.), stable hous. at T (miss.), employed at T (miss.); Called for a psy reason: age, years in France, years in France (miss.), months since CIR, months since CIR (miss.), months since protection (miss.), PHQ-2, prior contact, no educ. (miss.), higher educ., stable hous. (miss.), CIR hrs (miss.), stable hous. at T (miss.), employed at T (miss.); Helpline secured an appointment: age, years in France, years in France (miss.), months since CIR, months since CIR (miss.), months since protection (miss.), PHQ-2, prior contact, no educ., no educ. (miss.), higher educ., stable hous. (miss.), CIR hrs (miss.), stable hous. at T (miss.), employed at T, employed at T (miss.); Attended the helpline-secured appointment: age, years in France, years in France (miss.), months since CIR, months since CIR (miss.), months since protection (miss.), PHQ-2, prior contact, no educ., no educ. (miss.), higher educ., stable hous. (miss.), CIR hrs (miss.), stable hous. at T (miss.), employed at T (miss.).

D.2 Exit from the AGIR programme

Table D5: ITT Effects on Exit Outcomes

	Positive exit (housing, employment or CPH)			Other early exit			Reached the 24-month cap			Still enrolled at measurement date		
	(No cov.)	(With cov.)	(BCH-DL)	(No cov.)	(With cov.)	(BCH-DL)	(No cov.)	(With cov.)	(BCH-DL)	(No cov.)	(With cov.)	(BCH-DL)
Peer-Helper	-0.037	-0.039*	-0.041*	0.012	0.012	0.010	-0.002	0.001	0.003	0.027	0.027	0.032
	(0.024)	(0.022)	(0.022)	(0.018)	(0.018)	(0.018)	(0.041)	(0.029)	(0.029)	(0.042)	(0.031)	(0.031)
	[0.362]	[0.299]	[0.240]	[0.782]	[0.906]	[0.918]	[0.963]	[0.973]	[0.933]	[0.782]	[0.786]	[0.608]
Helpline	-0.021	-0.012	-0.012	-0.005	-0.003	-0.003	-0.074*	-0.043	-0.045	0.100**	0.058**	0.064**
	(0.025)	(0.024)	(0.024)	(0.016)	(0.017)	(0.016)	(0.040)	(0.029)	(0.029)	(0.042)	(0.029)	(0.029)
	[0.782]	[0.913]	[0.918]	[0.963]	[0.973]	[0.933]	[0.263]	[0.343]	[0.294]	[0.109]	[0.278]	[0.210]
Peer-Helper + Helpline	-0.054**	-0.045**	-0.046**	-0.001	-0.004	-0.005	0.004	0.001	0.002	0.051	0.048*	0.051*
	(0.023)	(0.022)	(0.022)	(0.017)	(0.017)	(0.017)	(0.041)	(0.028)	(0.028)	(0.042)	(0.029)	(0.029)
	[0.109]	[0.278]	[0.210]	[0.963]	[0.973]	[0.933]	[0.963]	[0.973]	[0.933]	[0.537]	[0.299]	[0.240]
Covariates	No	Yes	BCH-DL	No	Yes	BCH-DL	No	Yes	BCH-DL	No	Yes	BCH-DL
Control mean	0.107	0.107	0.107	0.041	0.041	0.041	0.390	0.390	0.390	0.462	0.462	0.462
Observations	1,135	1,135	1,135	1,135	1,135	1,135	1,135	1,135	1,135	1,135	1,135	1,135
Adjusted R^2	0.007	0.137	0.138	0.004	0.014	0.011	0.018	0.513	0.509	0.003	0.517	0.512
p: PH = Helpline	0.507	0.222	0.184	0.348	0.401	0.440	0.073	0.130	0.101	0.086	0.308	0.299
p: PH = PH + HL	0.402	0.781	0.827	0.457	0.354	0.381	0.889	0.989	0.980	0.564	0.478	0.523
p: Helpline = PH + HL	0.138	0.120	0.111	0.823	0.940	0.927	0.053	0.122	0.094	0.253	0.720	0.652
p: all jointly zero	0.103	0.116	0.101	0.812	0.798	0.830	0.155	0.345	0.279	0.111	0.204	0.149

Notes: Each column is a pooled three-arm OLS regression of the outcome on the three arm indicators with randomisation strata fixed effects (site x nationality group x gender). The three sub-columns per outcome are nested specifications: no covariates; the baseline covariate pool of the appendix on variable construction; and the subset of that pool selected by double-selection LASSO. Missing covariate values are recoded to a neutral placeholder with a paired missingness indicator, so no observation is dropped. Per arm: coefficient, HC1 robust standard error in parentheses, Benjamini-Hochberg q-value in square brackets computed within the specification. Footer: Control-arm mean, observations, pairwise tests of equality between arms, and the joint test that the three coefficients are zero. * p<0.10, ** p<0.05, *** p<0.01. BCH-DL retained the following covariates (per outcome): Positive exit (housing, employment or CPH): age, years in France, months in AGIR, months since CIR, months since CIR (miss.), months since protection, months since protection (miss.), PHQ-2, prior contact, higher educ., stable hous. (miss.), CIR hrs, CIR hrs (miss.), stable hous. at T, stable hous. at T (miss.), employed at T, employed at T (miss.); Other early exit: age, years in France, months since CIR, months since CIR (miss.), months since protection (miss.), PHQ-2, prior contact, hosp. visits BT, higher educ., stable hous. (miss.), stable hous. at T (miss.), employed at T, employed at T (miss.); Reached the 24-month cap: age, years in France, months in AGIR, months since CIR, months since CIR (miss.), months since protection (miss.), PHQ-2, prior contact, no educ. (miss.), higher educ., stable hous. (miss.), CIR hrs, stable hous. at T, stable hous. at T (miss.), employed at T, employed at T (miss.); Still enrolled at measurement date: age, years in France, years in France (miss.), months in AGIR, months since CIR, months since CIR (miss.), months since protection (miss.), PTSD (miss.), PHQ-2, prior contact, no educ. (miss.), higher educ., stable hous. (miss.), stable hous. at T, stable hous. at T (miss.), employed at T, employed at T (miss.).

D.3 Employment and housing trajectories

Table D6: ITT Effects on Trajectory (Employment and Housing) Outcomes

	Share of days active			Share of days employed			Share of days in stable housing		
	(No cov.)	(With cov.)	(BCH-DL)	(No cov.)	(With cov.)	(BCH-DL)	(No cov.)	(With cov.)	(BCH-DL)
Peer-Helper	-0.021	-0.034	-0.033	-0.021	-0.046	-0.046	-0.016	-0.018	-0.017
	(0.033)	(0.028)	(0.028)	(0.034)	(0.028)	(0.028)	(0.030)	(0.022)	(0.021)
	[0.771]	[0.718]	[0.749]	[0.771]	[0.718]	[0.749]	[0.771]	[0.718]	[0.749]
Helpline	-0.021	-0.018	-0.018	-0.017	-0.025	-0.024	-0.030	-0.001	-0.002
	(0.032)	(0.029)	(0.029)	(0.033)	(0.029)	(0.029)	(0.028)	(0.022)	(0.022)
	[0.771]	[0.718]	[0.749]	[0.771]	[0.718]	[0.749]	[0.771]	[0.998]	[0.988]
Peer-Helper + Helpline	0.007	-0.000	-0.000	-0.004	-0.016	-0.015	-0.031	-0.020	-0.018
	(0.032)	(0.028)	(0.028)	(0.033)	(0.028)	(0.028)	(0.028)	(0.021)	(0.021)
	[0.904]	[0.998]	[0.988]	[0.904]	[0.718]	[0.749]	[0.771]	[0.718]	[0.749]
Covariates	No	Yes	BCH-DL	No	Yes	BCH-DL	No	Yes	BCH-DL
Control mean	0.637	0.637	0.637	0.583	0.583	0.583	0.201	0.201	0.201
Observations	1,133	1,133	1,133	1,133	1,133	1,133	1,133	1,133	1,133
Adjusted R^2	0.130	0.358	0.358	0.152	0.397	0.397	0.116	0.497	0.495
p: PH = Helpline	0.998	0.580	0.607	0.918	0.478	0.437	0.637	0.471	0.517
p: PH = PH + HL	0.391	0.212	0.230	0.618	0.294	0.274	0.613	0.928	0.944
p: Helpline = PH + HL	0.379	0.520	0.523	0.684	0.753	0.770	0.977	0.399	0.454
p: all jointly zero	0.752	0.548	0.574	0.910	0.441	0.436	0.672	0.697	0.749

Notes: Each column is a pooled three-arm OLS regression of the outcome on the three arm indicators with randomisation strata fixed effects (site x nationality group x gender). The three sub-columns per outcome are nested specifications: no covariates; the baseline covariate pool of the appendix on variable construction; and the subset of that pool selected by double-selection LASSO. Missing covariate values are recoded to a neutral placeholder with a paired missingness indicator, so no observation is dropped. Per arm: coefficient, HC1 robust standard error in parentheses, Benjamini-Hochberg q-value in square brackets computed within the specification. Footer: Control-arm mean, observations, pairwise tests of equality between arms, and the joint test that the three coefficients are zero. * p<0.10, ** p<0.05, *** p<0.01. BCH-DL retained the following covariates (per outcome): Share of days active: share active BT, share employed BT, share stable hous. BT; Share of days employed: share active BT, share employed BT, share stable hous. BT, days AGIR to T; Share of days in stable housing: share stable hous. BT, months France to AGIR.

Table D7: ITT Effects on Trajectory Intensive Margins Outcomes

	Share of days in stable employment			Share of days in full-time employment			Share of days housed in own name		
	(No cov.)	(With cov.)	(BCH-DL)	(No cov.)	(With cov.)	(BCH-DL)	(No cov.)	(With cov.)	(BCH-DL)
Peer-Helper	-0.049	-0.053*	-0.054*	0.013	0.014	0.012	-0.005	0.000	0.000
	(0.036)	(0.029)	(0.029)	(0.036)	(0.028)	(0.028)	(0.022)	(0.015)	(0.015)
	[0.779]	[0.627]	[0.575]	[0.976]	[0.800]	[0.852]	[0.976]	[0.986]	[0.984]
Helpline	-0.060*	-0.039	-0.040	-0.001	0.001	0.001	0.001	0.020	0.017
	(0.035)	(0.030)	(0.029)	(0.034)	(0.028)	(0.028)	(0.022)	(0.016)	(0.016)
	[0.779]	[0.627]	[0.716]	[0.976]	[0.986]	[0.984]	[0.976]	[0.627]	[0.716]
Peer-Helper + Helpline	-0.034	-0.025	-0.024	0.015	0.018	0.020	-0.010	0.012	0.012
	(0.035)	(0.029)	(0.029)	(0.034)	(0.028)	(0.028)	(0.020)	(0.014)	(0.014)
	[0.976]	[0.745]	[0.716]	[0.976]	[0.778]	[0.716]	[0.976]	[0.745]	[0.716]
Covariates	No	Yes	BCH-DL	No	Yes	BCH-DL	No	Yes	BCH-DL
Control mean	0.473	0.473	0.473	0.396	0.396	0.396	0.098	0.098	0.098
Observations	1,133	1,133	1,133	1,133	1,133	1,133	1,133	1,133	1,133
Adjusted R^2	0.151	0.446	0.445	0.148	0.456	0.454	0.149	0.542	0.539
p: PH = Helpline	0.772	0.630	0.638	0.694	0.661	0.678	0.810	0.243	0.316
p: PH = PH + HL	0.676	0.331	0.291	0.942	0.880	0.781	0.809	0.457	0.466
p: Helpline = PH + HL	0.464	0.635	0.570	0.627	0.555	0.493	0.606	0.613	0.734
p: all jointly zero	0.347	0.315	0.286	0.947	0.894	0.872	0.950	0.550	0.634

Notes: Each column is a pooled three-arm OLS regression of the outcome on the three arm indicators with randomisation strata fixed effects (site x nationality group x gender). The three sub-columns per outcome are nested specifications: no covariates; the baseline covariate pool of the appendix on variable construction; and the subset of that pool selected by double-selection LASSO. Missing covariate values are recoded to a neutral placeholder with a paired missingness indicator, so no observation is dropped. Per arm: coefficient, HC1 robust standard error in parentheses, Benjamini-Hochberg q-value in square brackets computed within the specification. Footer: Control-arm mean, observations, pairwise tests of equality between arms, and the joint test that the three coefficients are zero. * p<0.10, ** p<0.05, *** p<0.01. BCH-DL retained the following covariates (per outcome): Share of days in stable employment: share stable emp. BT, share full-time BT, share own-name hous. BT, days AGIR to T; Share of days in full-time employment: share full-time BT, share own-name hous. BT, days AGIR to T; Share of days housed in own name: share stable emp. BT, share own-name hous. BT.

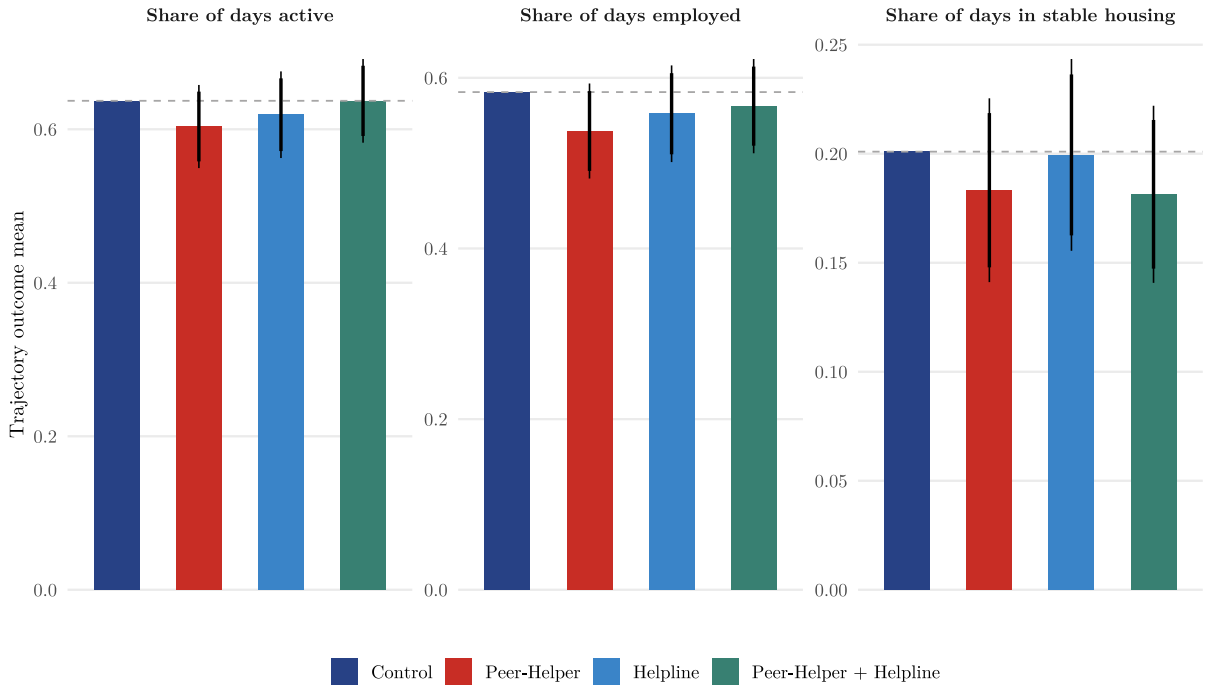

Figure D1: Effects on employment and housing trajectories, by treatment arm

Notes: Bars report the share of days spent in each state between the treatment date and the earlier of the AGIR exit and the measurement date, by treatment arm. The Control bar is the unadjusted Control-group mean; each treated-arm bar equals that mean plus the intention-to-treat estimate from the ANCOVA specification, with its 90% (thick) and 95% (thin) confidence intervals. The dashed horizontal line marks the Control mean.

D.4 Robustness

Table D8: ITT Effects on any-Signal (with Social-Worker FE) Outcomes

	Any post-treatment signal		
	(No cov.)	(With cov.)	(BCH-DL)
Peer-Helper	-0.017	-0.014	-0.014
	(0.022)	(0.022)	(0.022)
	[0.447]	[0.531]	[0.513]
Helpline	0.025	0.028	0.029
	(0.022)	(0.022)	(0.022)
	[0.385]	[0.312]	[0.281]
Peer-Helper + Helpline	0.062**	0.060**	0.062**
	(0.026)	(0.025)	(0.025)
	[0.045]	[0.053]	[0.037]
Covariates	No	Yes	BCH-DL
Control mean	0.062	0.062	0.062
Observations	1,132	1,132	1,132
Adjusted R^2	0.020	0.061	0.067
p: PH = Helpline	0.073	0.073	0.061
p: PH = PH + HL	0.003	0.004	0.002
p: Helpline = PH + HL	0.163	0.219	0.201
p: all jointly zero	0.016	0.019	0.011

Notes: Each column is a pooled three-arm OLS regression of the outcome on the three arm indicators with randomisation strata fixed effects (site x nationality group x gender). The three sub-columns per outcome are nested specifications: no covariates; the baseline covariate pool of the appendix on variable construction; and the subset of that pool selected by double-selection LASSO. Missing covariate values are recoded to a neutral placeholder with a paired missingness indicator, so no observation is dropped. Per arm: coefficient, HC1 robust standard error in parentheses, Benjamini-Hochberg q-value in square brackets computed within the specification. Footer: Control-arm mean, observations, pair-wise tests of equality between arms, and the joint test that the three coefficients are zero. * p<0.10, ** p<0.05, *** p<0.01. BCH-DL retained the following covariates (per outcome): Any post-treatment signal: age, months in AGIR, months since CIR, months since CIR (miss.), months since protection (miss.), PHQ-2, prior contact, hosp. visits BT, higher educ., stable hous. (miss.), stable hous. at T (miss.), employed at T, employed at T (miss.).

Table D9: ITT Effects on Hospital (with Social-Worker FE) Outcomes

	Any hospital visit (AP-HP)			Number of hospital visits (AP-HP)		
	(No cov.)	(With cov.)	(BCH-DL)	(No cov.)	(With cov.)	(BCH-DL)
Peer-Helper	-0.014	-0.013	-0.010	-0.005	-0.003	-0.001
	(0.020)	(0.020)	(0.020)	(0.025)	(0.024)	(0.025)
	[0.958]	[0.978]	[0.969]	[0.958]	[0.978]	[0.969]
Helpline	0.001	-0.004	-0.002	0.004	-0.001	0.001
	(0.019)	(0.019)	(0.019)	(0.020)	(0.021)	(0.021)
	[0.958]	[0.978]	[0.969]	[0.958]	[0.978]	[0.969]
Peer-Helper + Helpline	0.038	0.033	0.037	0.047*	0.041	0.045*
	(0.023)	(0.023)	(0.023)	(0.027)	(0.027)	(0.026)
	[0.316]	[0.466]	[0.334]	[0.316]	[0.466]	[0.334]
Covariates	No	Yes	BCH-DL	No	Yes	BCH-DL
Control mean	0.055	0.055	0.055	0.059	0.059	0.059
Observations	1,132	1,132	1,132	1,132	1,132	1,132
Adjusted R^2	0.018	0.016	0.023	0.006	0.001	0.009
p: PH = Helpline	0.466	0.643	0.674	0.700	0.906	0.936
p: PH = PH + HL	0.029	0.048	0.044	0.069	0.119	0.099
p: Helpline = PH + HL	0.108	0.107	0.094	0.102	0.111	0.092
p: all jointly zero	0.179	0.244	0.221	0.255	0.357	0.286

Notes: Each column is a pooled three-arm OLS regression of the outcome on the three arm indicators with randomisation strata fixed effects (site x nationality group x gender). The three sub-columns per outcome are nested specifications: no covariates; the baseline covariate pool of the appendix on variable construction; and the subset of that pool selected by double-selection LASSO. Missing covariate values are recoded to a neutral placeholder with a paired missingness indicator, so no observation is dropped. Per arm: coefficient, HC1 robust standard error in parentheses, Benjamini-Hochberg q-value in square brackets computed within the specification. Footer: Control-arm mean, observations, pairwise tests of equality between arms, and the joint test that the three coefficients are zero. * p<0.10, ** p<0.05, *** p<0.01. BCH-DL retained the following covariates (per outcome): Any hospital visit (AP-HP): age, months since CIR, months since CIR (miss.), months since protection (miss.), PHQ-2, prior contact, hosp. visits BT, higher educ., stable hous. (miss.), stable hous. at T (miss.), employed at T, employed at T (miss.); Number of hospital visits (AP-HP): age, months since CIR, months since CIR (miss.), months since protection (miss.), PHQ-2, prior contact, hosp. visits BT, higher educ., stable hous. (miss.), stable hous. at T (miss.), employed at T, employed at T (miss.).

Table D10: ITT Effects on Partner-Structure (with Social-Worker FE) Outcomes

	Any partner-structure appointment			Number of psy appointments (partner structures)			Number of appointments (partner structures, any type)		
	(No cov.)	(With cov.)	(BCH-DL)	(No cov.)	(With cov.)	(BCH-DL)	(No cov.)	(With cov.)	(BCH-DL)
Peer-Helper	0.000	-0.002	-0.002	-0.016	0.007	-0.002	-0.015	0.005	-0.004
	(0.008)	(0.008)	(0.007)	(0.045)	(0.027)	(0.029)	(0.045)	(0.028)	(0.029)
	[0.954]	[0.997]	[0.949]	[0.827]	[0.997]	[0.949]	[0.827]	[0.997]	[0.949]
Helpline	-0.005	0.005	0.005	-0.039	-0.000	-0.006	-0.044	0.000	-0.006
	(0.007)	(0.006)	(0.006)	(0.045)	(0.022)	(0.024)	(0.045)	(0.022)	(0.024)
	[0.827]	[0.997]	[0.949]	[0.827]	[0.997]	[0.949]	[0.827]	[0.997]	[0.949]
Peer-Helper + Helpline	0.012	0.014*	0.013*	-0.027	-0.013	-0.020	-0.024	-0.009	-0.017
	(0.010)	(0.008)	(0.008)	(0.044)	(0.033)	(0.037)	(0.045)	(0.033)	(0.037)
	[0.827]	[0.676]	[0.828]	[0.827]	[0.997]	[0.949]	[0.827]	[0.997]	[0.949]
Covariates	No	Yes	BCH-DL	No	Yes	BCH-DL	No	Yes	BCH-DL
Control mean	0.007	0.007	0.007	0.041	0.041	0.041	0.045	0.045	0.045
Observations	1,132	1,132	1,132	1,132	1,132	1,132	1,132	1,132	1,132
Adjusted R^2	-0.015	0.356	0.356	-0.006	0.189	0.180	-0.009	0.219	0.214
p: PH = Helpline	0.459	0.233	0.228	0.295	0.690	0.817	0.219	0.786	0.920
p: PH = PH + HL	0.306	0.074	0.071	0.592	0.439	0.451	0.723	0.604	0.622
p: Helpline = PH + HL	0.079	0.216	0.209	0.322	0.541	0.509	0.147	0.678	0.632
p: all jointly zero	0.349	0.222	0.247	0.612	0.894	0.900	0.375	0.965	0.963

Notes: Each column is a pooled three-arm OLS regression of the outcome on the three arm indicators with randomisation strata fixed effects (site x nationality group x gender). The three sub-columns per outcome are nested specifications: no covariates; the baseline covariate pool of the appendix on variable construction; and the subset of that pool selected by double-selection LASSO. Missing covariate values are recoded to a neutral placeholder with a paired missingness indicator, so no observation is dropped. Per arm: coefficient, HC1 robust standard error in parentheses, Benjamini-Hochberg q-value in square brackets computed within the specification. Footer: Control-arm mean, observations, pairwise tests of equality between arms, and the joint test that the three coefficients are zero. * p<0.10, ** p<0.05, *** p<0.01. BCH-DL retained the following covariates (per outcome): Any partner-structure appointment: age, months since CIR, months since CIR (miss.), months since protection (miss.), PHQ-2, prior contact, higher educ., stable hous. (miss.), stable hous. at T (miss.), employed at T (miss.); Number of psy appointments (partner structures): age, months since CIR, months since CIR (miss.), months since protection (miss.), PHQ-2, prior contact, higher educ., stable hous. (miss.), stable hous. at T (miss.), employed at T (miss.); Number of appointments (partner structures, any type): age, months since CIR, months since CIR (miss.), months since protection (miss.), PHQ-2, prior contact, higher educ., stable hous. (miss.), stable hous. at T (miss.), employed at T (miss.).

Table D11: ITT Effects on Helpline (with Social-Worker FE) Outcomes

	Called the helpline			Called for a psy reason			Helpline secured an appointment			Attended the helpline-secured appointment		
	(No cov.)	(With cov.)	(BCH-DL)	(No cov.)	(With cov.)	(BCH-DL)	(No cov.)	(With cov.)	(BCH-DL)	(No cov.)	(With cov.)	(BCH-DL)
Peer-Helper	-0.001	0.001	-0.000	-0.004	-0.002	-0.004	-0.002	-0.003	-0.004	-0.003	-0.004	-0.004
	(0.005)	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
	[0.780]	[0.819]	[0.958]	[0.499]	[0.819]	[0.422]	[0.481]	[0.410]	[0.320]	[0.475]	[0.312]	[0.287]
Helpline	0.034***	0.034***	0.034***	0.019**	0.021**	0.020**	0.007	0.009	0.008	0.004	0.004	0.005
	(0.011)	(0.011)	(0.011)	(0.009)	(0.010)	(0.009)	(0.006)	(0.006)	(0.006)	(0.004)	(0.005)	(0.005)
	[0.010]	[0.009]	[0.007]	[0.074]	[0.068]	[0.062]	[0.318]	[0.286]	[0.254]	[0.481]	[0.463]	[0.378]
Peer-Helper + Helpline	0.043***	0.045***	0.044***	0.034***	0.036***	0.035***	0.018**	0.019**	0.018**	0.015**	0.015**	0.015**
	(0.012)	(0.012)	(0.012)	(0.010)	(0.011)	(0.010)	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)
	[0.005]	[0.004]	[0.003]	[0.008]	[0.005]	[0.006]	[0.056]	[0.034]	[0.041]	[0.074]	[0.068]	[0.062]
Covariates	No	Yes	BCH-DL	No	Yes	BCH-DL	No	Yes	BCH-DL	No	Yes	BCH-DL
Control mean	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Observations	1,132	1,132	1,132	1,132	1,132	1,132	1,132	1,132	1,132	1,132	1,132	1,132
Adjusted R^2	0.039	0.071	0.072	0.053	0.071	0.072	-0.003	0.055	0.052	-0.006	0.030	0.028
p: PH = Helpline	0.002	0.002	0.001	0.022	0.017	0.013	0.164	0.087	0.081	0.280	0.174	0.151
p: PH = PH + HL	0.000	0.001	0.000	0.001	0.001	0.001	0.010	0.007	0.007	0.021	0.020	0.017
p: Helpline = PH + HL	0.561	0.489	0.513	0.281	0.269	0.272	0.255	0.258	0.277	0.179	0.205	0.202
p: all jointly zero	0.000	0.000	0.000	0.001	0.001	0.001	0.052	0.037	0.037	0.121	0.109	0.100

Notes: Each column is a pooled three-arm OLS regression of the outcome on the three arm indicators with randomisation strata fixed effects (site x nationality group x gender). The three sub-columns per outcome are nested specifications: no covariates; the baseline covariate pool of the appendix on variable construction; and the subset of that pool selected by double-selection LASSO. Missing covariate values are recoded to a neutral placeholder with a paired missingness indicator, so no observation is dropped. Per arm: coefficient, HC1 robust standard error in parentheses, Benjamini-Hochberg q-value in square brackets computed within the specification. Footer: Control-arm mean, observations, pairwise tests of equality between arms, and the joint test that the three coefficients are zero. * p<0.10, ** p<0.05, *** p<0.01. BCH-DL retained the following covariates (per outcome): Called the helpline: age, years in France (miss.), months since CIR, months since CIR (miss.), months since protection (miss.), PHQ-2, prior contact, hosp. visits BT, no educ. (miss.), higher educ., higher educ. (miss.), stable hous. (miss.), stable hous. at T (miss.), employed at T (miss.); Called for a psy reason: age, years in France (miss.), months since CIR, months since CIR (miss.), months since protection (miss.), PHQ-2, prior contact, no educ. (miss.), higher educ., stable hous. (miss.), CIR hrs (miss.), stable hous. at T (miss.), employed at T (miss.); Helpline secured an appointment: age, years in France (miss.), months since CIR, months since CIR (miss.), months since protection (miss.), PHQ-2, prior contact, no educ., no educ. (miss.), higher educ., stable hous. (miss.), CIR hrs (miss.), stable hous. at T (miss.), employed at T, employed at T (miss.); Attended the helpline-secured appointment: age, years in France (miss.), months since CIR, months since CIR (miss.), months since protection (miss.), PHQ-2, prior contact, no educ., no educ. (miss.), higher educ., stable hous. (miss.), CIR hrs (miss.), stable hous. at T (miss.), employed at T (miss.).

Table D12: ITT Effects on Exit (with Social-Worker FE) Outcomes

	Positive exit (housing, employment or CPH)			Other early exit			Reached the 24-month cap			Still enrolled at measurement date		
	(No cov.)	(With cov.)	(BCH-DL)	(No cov.)	(With cov.)	(BCH-DL)	(No cov.)	(With cov.)	(BCH-DL)	(No cov.)	(With cov.)	(BCH-DL)
Peer-Helper	-0.053**	-0.054**	-0.057**	0.015	0.012	0.013	-0.002	0.018	0.019	0.040	0.024	0.028
	(0.025)	(0.024)	(0.024)	(0.019)	(0.019)	(0.019)	(0.037)	(0.029)	(0.029)	(0.037)	(0.031)	(0.030)
	[0.098]	[0.104]	[0.070]	[0.637]	[0.709]	[0.672]	[0.947]	[0.709]	[0.672]	[0.470]	[0.709]	[0.606]
Helpline	-0.036	-0.027	-0.027	-0.004	-0.005	-0.003	-0.058	-0.031	-0.031	0.099***	0.063**	0.067**
	(0.026)	(0.025)	(0.025)	(0.017)	(0.018)	(0.017)	(0.037)	(0.029)	(0.029)	(0.036)	(0.028)	(0.028)
	[0.317]	[0.579]	[0.556]	[0.947]	[0.897]	[0.926]	[0.273]	[0.579]	[0.556]	[0.041]	[0.104]	[0.070]
Peer-Helper + Helpline	-0.075***	-0.061***	-0.062***	0.003	-0.002	-0.001	-0.005	0.006	0.008	0.077**	0.057**	0.060**
	(0.024)	(0.023)	(0.023)	(0.018)	(0.018)	(0.018)	(0.036)	(0.028)	(0.028)	(0.035)	(0.027)	(0.027)
	[0.020]	[0.102]	[0.070]	[0.947]	[0.897]	[0.958]	[0.947]	[0.897]	[0.926]	[0.098]	[0.113]	[0.086]
Covariates	No	Yes	BCH-DL	No	Yes	BCH-DL	No	Yes	BCH-DL	No	Yes	BCH-DL
Control mean	0.107	0.107	0.107	0.041	0.041	0.041	0.390	0.390	0.390	0.462	0.462	0.462
Observations	1,132	1,132	1,132	1,132	1,132	1,132	1,132	1,132	1,132	1,132	1,132	1,132
Adjusted R^2	0.033	0.156	0.156	0.004	0.015	0.010	0.297	0.574	0.572	0.327	0.589	0.588
p: PH = Helpline	0.475	0.218	0.177	0.291	0.324	0.359	0.113	0.083	0.075	0.106	0.190	0.186
p: PH = PH + HL	0.297	0.743	0.787	0.482	0.397	0.420	0.946	0.644	0.678	0.296	0.244	0.264
p: Helpline = PH + HL	0.085	0.116	0.104	0.684	0.862	0.883	0.122	0.181	0.152	0.536	0.839	0.785
p: all jointly zero	0.016	0.040	0.030	0.756	0.774	0.806	0.299	0.355	0.324	0.034	0.084	0.062

Notes: Each column is a pooled three-arm OLS regression of the outcome on the three arm indicators with randomisation strata fixed effects (site x nationality group x gender). The three sub-columns per outcome are nested specifications: no covariates; the baseline covariate pool of the appendix on variable construction; and the subset of that pool selected by double-selection LASSO. Missing covariate values are recoded to a neutral placeholder with a paired missingness indicator, so no observation is dropped. Per arm: coefficient, HC1 robust standard error in parentheses, Benjamini-Hochberg q-value in square brackets computed within the specification. Footer: Control-arm mean, observations, pairwise tests of equality between arms, and the joint test that the three coefficients are zero. * p<0.10, ** p<0.05, *** p<0.01. BCH-DL retained the following covariates (per outcome): Positive exit (housing, employment or CPH): age, years in France, months in AGIR, months since CIR, months since CIR (miss.), months since protection, months since protection (miss.), PHQ-2, prior contact, higher educ., stable hous. (miss.), CIR hrs, CIR hrs (miss.), stable hous. at T, stable hous. at T (miss.), employed at T, employed at T (miss.); Other early exit: age, months since CIR, months since CIR (miss.), months since protection (miss.), PHQ-2, prior contact, hosp. visits BT, higher educ., stable hous. (miss.), stable hous. at T (miss.), employed at T, employed at T (miss.); Reached the 24-month cap: age, months in AGIR, months since CIR, months since CIR (miss.), months since protection (miss.), PHQ-2, prior contact, no educ. (miss.), higher educ., stable hous. (miss.), CIR hrs, stable hous. at T, stable hous. at T (miss.), employed at T, employed at T (miss.); Still enrolled at measurement date: age, years in France (miss.), months in AGIR, months since CIR, months since CIR (miss.), months since protection (miss.), PTSD (miss.), PHQ-2, prior contact, no educ. (miss.), higher educ., stable hous. (miss.), stable hous. at T, stable hous. at T (miss.), employed at T, employed at T (miss.).

D.5 Effects on the Treated

Table D13: 2SLS estimates: effect of any post-treatment healthcare contact on trajectory outcomes

Outcome	Control mean	(No cov.)	(With cov.)
Share of days active	0.637	0.203	0.231
		(0.350)	(0.300)
		[0.814]	[0.690]
Share of days employed	0.583	0.085	0.120
		(0.360)	(0.302)
		[0.814]	[0.690]
Share of days in stable housing	0.201	-0.286	-0.097
		(0.308)	(0.224)
		[0.814]	[0.690]
Covariates		No	Yes
First-stage F (Kleibergen-Paap)		4.6	4.5
Observations		1,133	1,133

Notes: Each cell of the top panel reports the 2SLS coefficient on an indicator for any post-treatment healthcare contact (at least one partner-structure appointment, hospital admission or helpline call after the treatment date), instrumented by the three random assignment dummies on the full four-arm sample, with strata fixed effects (site x nationality group x gender) and HC1 heteroskedasticity-robust standard errors in parentheses. Outcomes are the share of days spent in each state between the treatment date and AGIR exit or the measurement date, whichever comes first. The (With cov.) column adds the baseline trajectory shares and the programme-timing controls of the ANCOVA specification. A low first-stage F-statistic signals a weak-instrument problem and confidence intervals that understate the true uncertainty; the estimates rescale a near-zero reduced form and should be read as bounds on the complier effect. Benjamini-Hochberg q-values in square brackets, pooled across the outcomes of each column. * p<0.10, ** p<0.05, *** p<0.01.

Table D14: 2SLS estimates: effect of any post-treatment healthcare contact on intensive-margin trajectory outcomes

Outcome	Control mean	(No cov.)	(With cov.)
Share of days in stable employment	0.473	-0.122	0.045
		(0.387)	(0.307)
		[0.833]	[0.884]
Share of days in full-time employment	0.396	0.079	0.108
		(0.373)	(0.298)
		[0.833]	[0.884]
Share of days housed in own name	0.098	-0.077	0.169
		(0.223)	(0.173)
		[0.833]	[0.884]
Covariates		No	Yes
First-stage F (Kleibergen-Paap)		4.6	4.5
Observations		1,133	1,133

Notes: Each cell of the top panel reports the 2SLS coefficient on an indicator for any post-treatment healthcare contact (at least one partner-structure appointment, hospital admission or helpline call after the treatment date), instrumented by the three random assignment dummies on the full four-arm sample, with strata fixed effects (site x nationality group x gender) and HC1 heteroskedasticity-robust standard errors in parentheses. Outcomes are the share of days spent in each state between the treatment date and AGIR exit or the measurement date, whichever comes first. The (With cov.) column adds the baseline intensive-margin shares (stable employment, full-time employment, own-name housing) and the programme-timing controls. A low first-stage F-statistic signals a weak-instrument problem and confidence intervals that understate the true uncertainty; the estimates rescale a near-zero reduced form and should be read as bounds on the complier effect. Benjamini-Hochberg q-values in square brackets, pooled across the outcomes of each column. * p<0.10, ** p<0.05, *** p<0.01.

Table D15: 2SLS estimates: effect of calling the helpline on trajectory outcomes (LATE)

		Helpline		Peer-Helper + Helpline	
Outcome	Control mean	(No cov.)	(With cov.)	(No cov.)	(With cov.)
Share of days active	0.637	-0.522	-0.560	0.264	0.037
		(1.042)	(0.967)	(0.751)	(0.667)
		[0.701]	[0.844]	[0.975]	[0.955]
Share of days employed	0.583	-0.410	-0.725	0.024	-0.310
		(1.067)	(0.983)	(0.776)	(0.686)
		[0.701]	[0.844]	[0.975]	[0.955]
Share of days in stable housing	0.201	-0.819	0.015	-0.733	-0.406
		(0.940)	(0.741)	(0.681)	(0.499)
		[0.701]	[0.984]	[0.846]	[0.955]
Covariates		No	Yes	No	Yes
First-stage F (Kleibergen-Paap)		9.2	9.1	13.0	12.7
Take-up rate		0.033	0.033	0.046	0.046
Observations		563	563	575	575

Notes: Each cell of the top panel reports the 2SLS coefficient on an indicator for having called the helpline at least once after the treatment date, instrumented by the random assignment dummy of the column's helpline-access arm on the Control + arm sub-sample, with strata fixed effects (site x nationality group x gender) and HC1 heteroskedasticity-robust standard errors in parentheses. Outcomes are the share of days spent in each state between the treatment date and AGIR exit or the measurement date, whichever comes first. The (With cov.) columns add the baseline trajectory shares and the programme-timing controls of the ANCOVA specification. Compliance is one-sided (no Control refugee receives the number), so the first-stage coefficient equals the arm take-up rate; a low first-stage F-statistic signals a weak instrument and confidence intervals that understate the true uncertainty. In the combined arm, the peer-helper meeting may affect trajectories through channels other than the call, so that column's estimates should not be read as a pure helpline effect. Benjamini-Hochberg q-values in square brackets, pooled across the outcomes of each column. * p<0.10, ** p<0.05, *** p<0.01.

Table D16: 2SLS estimates: effect of calling the helpline on intensive-margin trajectory outcomes (LATE)

		Helpline		Peer-Helper + Helpline	
Outcome	Control mean	(No cov.)	(With cov.)	(No cov.)	(With cov.)
Share of days in stable employment	0.473	-1.745	-1.177	-0.610	-0.431
		(1.259)	(1.089)	(0.851)	(0.711)
		[0.499]	[0.420]	[0.626]	[0.545]
Share of days in full-time employment	0.396	0.148	0.045	0.489	0.510
		(1.082)	(0.955)	(0.815)	(0.677)
		[0.892]	[0.962]	[0.626]	[0.545]
Share of days housed in own name	0.098	0.109	0.759	-0.231	0.373
		(0.690)	(0.595)	(0.473)	(0.350)
		[0.892]	[0.420]	[0.626]	[0.545]
Covariates		No	Yes	No	Yes
First-stage F (Kleibergen-Paap)		9.2	9.0	13.0	12.8
Take-up rate		0.033	0.033	0.046	0.046
Observations		563	563	575	575

Notes: Each cell of the top panel reports the 2SLS coefficient on an indicator for having called the helpline at least once after the treatment date, instrumented by the random assignment dummy of the column's helpline-access arm on the Control + arm sub-sample, with strata fixed effects (site x nationality group x gender) and HC1 heteroskedasticity-robust standard errors in parentheses. Outcomes are the share of days spent in each state between the treatment date and AGIR exit or the measurement date, whichever comes first. The (With cov.) columns add the baseline intensive-margin shares (stable employment, full-time employment, own-name housing) and the programme-timing controls. Compliance is one-sided (no Control refugee receives the number), so the first-stage coefficient equals the arm take-up rate; a low first-stage F-statistic signals a weak instrument and confidence intervals that understate the true uncertainty. In the combined arm, the peer-helper meeting may affect trajectories through channels other than the call, so that column's estimates should not be read as a pure helpline effect. Benjamini-Hochberg q-values in square brackets, pooled across the outcomes of each column. * p<0.10, ** p<0.05, *** p<0.01.

Table D17: 2SLS estimates: effect of any post-treatment healthcare contact on AGIR exit outcomes

Outcome	Control mean	(No cov.)	(With cov.)
Still enrolled at measurement date	0.462	0.562	0.490
		(0.471)	(0.360)
		[0.466]	[0.576]
Reached the 24-month cap	0.390	-0.071	-0.107
		(0.435)	(0.314)
		[0.870]	[0.734]
Positive exit (housing, employment or CPH)	0.107	-0.380	-0.235
		(0.247)	(0.238)
		[0.466]	[0.576]
Other early exit	0.041	-0.111	-0.148
		(0.181)	(0.188)
		[0.721]	[0.576]
Covariates		No	Yes
First-stage F (Kleibergen-Paap)		4.6	4.7
Observations		1,135	1,135

Notes: Each cell of the top panel reports the 2SLS coefficient on an indicator for any post-treatment healthcare contact (at least one partner-structure appointment, hospital admission or helpline call after the treatment date), instrumented by the three random assignment dummies on the full four-arm sample, with strata fixed effects (site x nationality group x gender) and HC1 heteroskedasticity-robust standard errors in parentheses. Outcomes are the share of days spent in each state between the treatment date and AGIR exit or the measurement date, whichever comes first. The (With cov.) column adds the baseline covariate pool of the ITT exit table. A low first-stage F-statistic signals a weak-instrument problem and confidence intervals that understate the true uncertainty; the estimates rescale a near-zero reduced form and should be read as bounds on the complier effect. Benjamini-Hochberg q-values in square brackets, pooled across the outcomes of each column. * p<0.10, ** p<0.05, *** p<0.01.

E Heterogeneous Effects

E.1 Effects by Gender

Table E1: Heterogeneous Effects by Gender: Any Post-Treatment Signal

	Any post-treatment signal		
	Men	Women	p diff
Peer-Helper	0.001	-0.061	0.191
	(0.023)	(0.042)	
	[0.973]	[0.210]	
Helpline	0.015	0.106*	0.174
	(0.024)	(0.063)	
	[0.642]	[0.183]	
Peer-Helper + Helpline	0.059**	0.155**	0.211
	(0.026)	(0.072)	
	[0.093]	[0.093]	
Control mean	0.060	0.070	
Observations	936	199	
p: Peer-Helper = Helpline	0.567	0.004	
p: Peer-Helper = PH + Helpline	0.026	0.001	
p: Helpline = PH + Helpline	0.102	0.562	
p: all interactions zero			0.027

Notes: For each outcome and arm, the table reports the ITT effect for men (arm coefficient), the effect for women (arm coefficient plus its interaction with the female indicator, delta-method standard error) and the p-value of the male-female difference, from one OLS regression per outcome of the outcome on the three arm indicators fully interacted with the female indicator, with randomisation strata fixed effects and HC1 robust standard errors in parentheses. No covariates beyond the strata fixed effects. The female main effect is absorbed by the strata. Benjamini-Hochberg q-values in square brackets. Footer: Control-arm means and observations by sex, pairwise tests of equality between arms by sex, and the joint test that the three interactions are zero. A dash marks a cell that cannot be estimated because the outcome has no variation within that sex. * p<0.10, ** p<0.05, *** p<0.01.

Table E2: Heterogeneous Effects by Gender: Hospital Outcomes

	Any hospital visit (AP-HP)			Number of hospital visits (AP-HP)		
	Men	Women	p diff	Men	Women	p diff
Peer-Helper	0.003	-0.076**	0.059	0.010	-0.076**	0.054
	(0.021)	(0.036)		(0.027)	(0.036)	
	[0.882]	[0.186]		[0.767]	[0.186]	
Helpline	-0.011	0.092	0.112	-0.010	0.092	0.120
	(0.020)	(0.062)		(0.022)	(0.062)	
	[0.767]	[0.269]		[0.767]	[0.269]	
Peer-Helper + Helpline	0.029	0.129*	0.169	0.037	0.129*	0.216
	(0.022)	(0.069)		(0.027)	(0.069)	
	[0.295]	[0.186]		[0.270]	[0.186]	
Control mean	0.052	0.070		0.056	0.070	
Observations	936	199		936	199	
p: Peer-Helper = Helpline	0.497	0.002		0.438	0.002	
p: Peer-Helper = PH + Helpline	0.269	0.001		0.363	0.001	
p: Helpline = PH + Helpline	0.068	0.643		0.067	0.643	
p: all interactions zero			0.001			0.001

Notes: For each outcome and arm, the table reports the ITT effect for men (arm coefficient), the effect for women (arm coefficient plus its interaction with the female indicator, delta-method standard error) and the p-value of the male-female difference, from one OLS regression per outcome of the outcome on the three arm indicators fully interacted with the female indicator, with randomisation strata fixed effects and HC1 robust standard errors in parentheses. No covariates beyond the strata fixed effects. The female main effect is absorbed by the strata. Benjamini-Hochberg q-values in square brackets. Footer: Control-arm means and observations by sex, pairwise tests of equality between arms by sex, and the joint test that the three interactions are zero. A dash marks a cell that cannot be estimated because the outcome has no variation within that sex. * p<0.10, ** p<0.05, *** p<0.01.

Table E3: Heterogeneous Effects by Gender: Hospital Outcomes (with covariates)

	Any hospital visit (AP-HP)			Number of hospital visits (AP-HP)		
	Men	Women	p diff	Men	Women	p diff
Peer-Helper	-0.001	-0.063*	0.138	0.008	-0.064*	0.108
	(0.021)	(0.036)		(0.026)	(0.037)	
	[0.975]	[0.253]		[0.814]	[0.253]	
Helpline	-0.015	0.087	0.120	-0.013	0.084	0.146
	(0.020)	(0.062)		(0.022)	(0.063)	
	[0.624]	[0.364]		[0.684]	[0.364]	
Peer-Helper + Helpline	0.023	0.136*	0.126	0.033	0.130*	0.199
	(0.023)	(0.070)		(0.027)	(0.071)	
	[0.460]	[0.253]		[0.375]	[0.253]	
Control mean	0.052	0.070		0.056	0.070	
Observations	936	199		936	199	
p: Peer-Helper = Helpline	0.508	0.008		0.423	0.010	
p: Peer-Helper = PH + Helpline	0.313	0.001		0.428	0.002	
p: Helpline = PH + Helpline	0.085	0.543		0.089	0.571	
p: all interactions zero			0.005			0.010

Notes: For each outcome and arm, the table reports the ITT effect for men (arm coefficient), the effect for women (arm coefficient plus its interaction with the female indicator, delta-method standard error) and the p-value of the male-female difference, from one OLS regression per outcome of the outcome on the three arm indicators fully interacted with the female indicator, with randomisation strata fixed effects and HC1 robust standard errors in parentheses. The baseline covariate pool of the corresponding ITT table is included. The female main effect is absorbed by the strata. Benjamini-Hochberg q-values in square brackets. Footer: Control-arm means and observations by sex, pairwise tests of equality between arms by sex, and the joint test that the three interactions are zero. A dash marks a cell that cannot be estimated because the outcome has no variation within that sex. * p<0.10, ** p<0.05, *** p<0.01.

Table E4: Heterogeneous Effects by Gender: Partner-Structure Outcomes

	Any partner-structure appointment			Number of psy appointments (partner structures)			Number of appointments (partner structures, any type)		
	Men	Women	p diff	Men	Women	p diff	Men	Women	p diff
Peer-Helper	-0.001	0.018	0.410	-0.031	–	–	-0.031	0.021	0.398
	(0.008)	(0.021)		(0.057)			(0.058)	(0.021)	
	[0.904]	[0.741]		[0.741]			[0.741]	[0.741]	
Helpline	-0.005	-0.003	0.791	-0.050	–	–	-0.055	0.001	0.312
	(0.007)	(0.005)		(0.055)			(0.055)	(0.002)	
	[0.741]	[0.741]		[0.741]			[0.741]	[0.845]	
Peer-Helper + Helpline	0.011	0.025	0.577	-0.036	–	–	-0.028	0.000	0.598
	(0.010)	(0.024)		(0.053)			(0.054)	(0.002)	
	[0.741]	[0.741]		[0.741]			[0.741]	[0.902]	
Control mean	0.009	0.000		0.052	0.000		0.056	0.000	
Observations	936	199		936	199		936	199	
p: Peer-Helper = Helpline	0.561	0.317		0.384	–		0.295	0.313	
p: Peer-Helper = PH + Helpline	0.248	0.816		0.830	–		0.911	0.313	
p: Helpline = PH + Helpline	0.101	0.293		0.256	–		0.085	0.900	
p: all interactions zero			0.795			0.544			0.331

Notes: For each outcome and arm, the table reports the ITT effect for men (arm coefficient), the effect for women (arm coefficient plus its interaction with the female indicator, delta-method standard error) and the p-value of the male-female difference, from one OLS regression per outcome of the outcome on the three arm indicators fully interacted with the female indicator, with randomisation strata fixed effects and HC1 robust standard errors in parentheses. No covariates beyond the strata fixed effects. The female main effect is absorbed by the strata. Benjamini-Hochberg q-values in square brackets. Footer: Control-arm means and observations by sex, pairwise tests of equality between arms by sex, and the joint test that the three interactions are zero. A dash marks a cell that cannot be estimated because the outcome has no variation within that sex. * p<0.10, ** p<0.05, *** p<0.01.

Table E5: Heterogeneous Effects by Gender: Helpline Outcomes

	Called the helpline			Called for a psy reason			Helpline secured an appointment			Attended the helpline-secured appointment		
	Men	Women	p diff	Men	Women	p diff	Men	Women	p diff	Men	Women	p diff
Peer-Helper	-0.002	-0.006	0.577	-0.002	-0.006	0.618	-0.001	-0.003	0.702	-0.001	–	–
	(0.002)	(0.008)		(0.002)	(0.008)		(0.002)	(0.005)		(0.002)		
	[0.580]	[0.580]		[0.580]	[0.580]		[0.580]	[0.580]		[0.615]		
Helpline	0.034***	0.014	0.321	0.016*	0.014	0.898	0.004	0.017	0.442	0.004	–	–
	(0.012)	(0.017)		(0.009)	(0.017)		(0.004)	(0.016)		(0.004)		
	[0.034]	[0.580]		[0.209]	[0.580]		[0.580]	[0.580]		[0.580]		
Peer-Helper + Helpline	0.046***	0.025	0.422	0.038***	0.025	0.624	0.023**	0.000	0.023	0.020**	–	–
	(0.013)	(0.022)		(0.012)	(0.022)		(0.010)	(0.003)		(0.009)		
	[0.007]	[0.580]		[0.015]	[0.580]		[0.076]	[1.000]		[0.111]		
Control mean	0.000	0.000		0.000	0.000		0.000	0.000		0.000	0.000	
Observations	936	199		936	199		936	199		936	199	
p: Peer-Helper = Helpline	0.006	0.305		0.060	0.305		0.367	0.297		0.341	–	
p: Peer-Helper = PH + Helpline	0.000	0.215		0.001	0.215		0.011	0.489		0.019	1.000	
p: Helpline = PH + Helpline	0.523	0.696		0.161	0.696		0.076	0.299		0.127	1.000	
p: all interactions zero			0.471			0.882			0.108			0.108

Notes: For each outcome and arm, the table reports the ITT effect for men (arm coefficient), the effect for women (arm coefficient plus its interaction with the female indicator, delta-method standard error) and the p-value of the male-female difference, from one OLS regression per outcome of the outcome on the three arm indicators fully interacted with the female indicator, with randomisation strata fixed effects and HC1 robust standard errors in parentheses. No covariates beyond the strata fixed effects. The female main effect is absorbed by the strata. Benjamini-Hochberg q-values in square brackets. Footer: Control-arm means and observations by sex, pairwise tests of equality between arms by sex, and the joint test that the three interactions are zero. A dash marks a cell that cannot be estimated because the outcome has no variation within that sex. * p<0.10, ** p<0.05, *** p<0.01.

Table E6: Heterogeneous Effects by Gender: AGIR Exit Outcomes

	Positive exit (housing, employment or CPH)			Other early exit			Reached the 24-month cap			Still enrolled at measurement date		
	Men	Women	p diff	Men	Women	p diff	Men	Women	p diff	Men	Women	p diff
Peer-Helper	-0.053**	0.038	0.122	0.006	0.038	0.575	0.000	-0.010	0.921	0.047	-0.066	0.294
	(0.027)	(0.053)		(0.018)	(0.055)		(0.045)	(0.094)		(0.046)	(0.097)	
	[0.357]	[0.917]		[0.992]	[0.917]		[0.993]	[0.993]		[0.745]	[0.917]	
Helpline	-0.031	0.019	0.389	-0.008	0.010	0.725	-0.057	-0.148*	0.358	0.096**	0.118	0.831
	(0.029)	(0.050)		(0.016)	(0.051)		(0.045)	(0.088)		(0.047)	(0.095)	
	[0.745]	[0.987]		[0.985]	[0.993]		[0.661]	[0.561]		[0.357]	[0.661]	
Peer-Helper + Helpline	-0.057**	-0.050	0.896	-0.007	0.029	0.551	0.006	-0.002	0.939	0.057	0.023	0.756
	(0.026)	(0.041)		(0.016)	(0.059)		(0.045)	(0.095)		(0.046)	(0.102)	
	[0.357]	[0.661]		[0.985]	[0.985]		[0.993]	[0.993]		[0.661]	[0.993]	
Control mean	0.116	0.070		0.034	0.070		0.395	0.368		0.455	0.491	
Observations	936	199		936	199		936	199		936	199	
p: Peer-Helper = Helpline	0.379	0.747		0.407	0.641		0.202	0.116		0.300	0.058	
p: Peer-Helper = PH + Helpline	0.878	0.065		0.438	0.886		0.892	0.931		0.822	0.394	
p: Helpline = PH + Helpline	0.294	0.118		0.947	0.760		0.154	0.109		0.412	0.351	
p: all interactions zero			0.330			0.925			0.762			0.615

Notes: For each outcome and arm, the table reports the ITT effect for men (arm coefficient), the effect for women (arm coefficient plus its interaction with the female indicator, delta-method standard error) and the p-value of the male-female difference, from one OLS regression per outcome of the outcome on the three arm indicators fully interacted with the female indicator, with randomisation strata fixed effects and HC1 robust standard errors in parentheses. No covariates beyond the strata fixed effects. The female main effect is absorbed by the strata. Benjamini-Hochberg q-values in square brackets. Footer: Control-arm means and observations by sex, pairwise tests of equality between arms by sex, and the joint test that the three interactions are zero. A dash marks a cell that cannot be estimated because the outcome has no variation within that sex. * p<0.10, ** p<0.05, *** p<0.01.

Table E7: Heterogeneous Effects by Gender: Trajectory Outcomes

	Share of days active			Share of days employed			Share of days in stable housing		
	Men	Women	p diff	Men	Women	p diff	Men	Women	p diff
Peer-Helper	-0.041	0.070	0.201	-0.036	0.052	0.309	-0.027	0.032	0.524
	(0.036)	(0.079)		(0.038)	(0.077)		(0.031)	(0.088)	
	[0.821]	[0.821]		[0.821]	[0.821]		[0.821]	[0.956]	
Helpline	-0.025	-0.005	0.821	-0.008	-0.055	0.594	-0.036	-0.006	0.715
	(0.035)	(0.083)		(0.036)	(0.079)		(0.030)	(0.075)	
	[0.821]	[0.956]		[0.956]	[0.821]		[0.821]	[0.956]	
Peer-Helper + Helpline	0.003	0.015	0.899	0.010	-0.085	0.257	-0.049*	0.061	0.205
	(0.034)	(0.081)		(0.036)	(0.076)		(0.029)	(0.081)	
	[0.956]	[0.956]		[0.956]	[0.821]		[0.821]	[0.821]	
Control mean	0.690	0.419		0.643	0.340		0.172	0.319	
Observations	935	198		935	198		935	198	
p: Peer-Helper = Helpline	0.659	0.358		0.469	0.196		0.776	0.652	
p: Peer-Helper = PH + Helpline	0.211	0.480		0.219	0.082		0.466	0.755	
p: Helpline = PH + Helpline	0.402	0.820		0.597	0.719		0.651	0.408	
p: all interactions zero			0.551			0.190			0.631

Notes: For each outcome and arm, the table reports the ITT effect for men (arm coefficient), the effect for women (arm coefficient plus its interaction with the female indicator, delta-method standard error) and the p-value of the male-female difference, from one OLS regression per outcome of the outcome on the three arm indicators fully interacted with the female indicator, with randomisation strata fixed effects and HC1 robust standard errors in parentheses. No covariates beyond the strata fixed effects. The female main effect is absorbed by the strata. Benjamini-Hochberg q-values in square brackets. Footer: Control-arm means and observations by sex, pairwise tests of equality between arms by sex, and the joint test that the three interactions are zero. A dash marks a cell that cannot be estimated because the outcome has no variation within that sex. * p<0.10, ** p<0.05, *** p<0.01.

Table E8: Heterogeneous Effects by Gender: Intensive-Margin Trajectory Outcomes

	Share of days in stable employment			Share of days in full-time employment			Share of days housed in own name		
	Men	Women	p diff	Men	Women	p diff	Men	Women	p diff
Peer-Helper	-0.060	0.000	0.459	-0.008	0.110*	0.115	-0.001	-0.021	0.784
	(0.041)	(0.070)		(0.041)	(0.063)		(0.022)	(0.072)	
	[0.881]	[0.998]		[0.955]	[0.881]		[0.998]	[0.955]	
Helpline	-0.065	-0.038	0.750	-0.008	0.028	0.593	0.012	-0.051	0.325
	(0.039)	(0.073)		(0.040)	(0.054)		(0.023)	(0.060)	
	[0.881]	[0.955]		[0.955]	[0.955]		[0.955]	[0.955]	
Peer-Helper + Helpline	-0.036	-0.031	0.958	0.019	-0.016	0.598	-0.014	0.018	0.651
	(0.040)	(0.071)		(0.039)	(0.054)		(0.020)	(0.067)	
	[0.955]	[0.955]		[0.955]	[0.955]		[0.955]	[0.955]	
Control mean	0.526	0.257		0.472	0.088		0.075	0.193	
Observations	935	198		935	198		935	198	
p: Peer-Helper = Helpline	0.909	0.611		0.993	0.202		0.573	0.650	
p: Peer-Helper = PH + Helpline	0.554	0.671		0.498	0.044		0.555	0.579	
p: Helpline = PH + Helpline	0.458	0.931		0.492	0.424		0.230	0.256	
p: all interactions zero			0.883			0.206			0.505

Notes: For each outcome and arm, the table reports the ITT effect for men (arm coefficient), the effect for women (arm coefficient plus its interaction with the female indicator, delta-method standard error) and the p-value of the male-female difference, from one OLS regression per outcome of the outcome on the three arm indicators fully interacted with the female indicator, with randomisation strata fixed effects and HC1 robust standard errors in parentheses. No covariates beyond the strata fixed effects. The female main effect is absorbed by the strata. Benjamini-Hochberg q-values in square brackets. Footer: Control-arm means and observations by sex, pairwise tests of equality between arms by sex, and the joint test that the three interactions are zero. A dash marks a cell that cannot be estimated because the outcome has no variation within that sex. * p<0.10, ** p<0.05, *** p<0.01.

E.2 Counterfactual Peer Assignment

Peer helpers were assigned to treated refugees by intervention language and gender, following a fixed priority order among the mediators present on the treatment day. Because assignment to arms is random and the mediator pool available on a given half-day is fixed, the mediator a Control refugee *would have* received is a deterministic function of pre-treatment characteristics and can be reconstructed. The first constraint is availability: only mediators present on the refugee's treatment half-day qualify, the half-day being taken from the start time of the baseline questionnaire administered in the treatment session. The second constraint is language, through priority lists defined by the refugee's intervention language and gender; ties and residual cases are resolved by a seeded random draw. Refugees already assigned to a peer arm keep their actual mediator.

Table E9 validates the reconstruction on refugees actually assigned to a peer arm, and Table E10 reports the resulting distribution of actual and counterfactual mediators. Table E11 reports the peer-assignment effect within each mediator cell, pooling the two peer arms; the arm-specific version is discussed in the body of the paper.

Table E9: Counterfactual peer assignment: validation on treated refugees

Metric	Value
Treated refugees with a listed mediator	557
Treatment days covered by the schedule	99.3%
Would-be language predictor matches the actual intervention language	84.6%
Actual mediator belongs to the priority set for the language and gender	84.2%
Algorithm's first available choice matches the actual mediator	72.9%

Notes: Validation of the counterfactual peer-assignment algorithm on refugees actually assigned to a peer arm. The would-be intervention language is the modal intervention language among treated refugees with the same mother tongue; the priority set is the ordered list of eligible mediators for the refugee's language and gender; the first available choice further restricts that list to mediators present on the refugee's treatment day according to the schedule, breaking ties at random with a fixed seed. Presence is matched at the half-day level, using the start time of the baseline questionnaire administered in the treatment session; the few refugees without a recorded time are matched at the day level.

Table E10: Actual (peer arms) and counterfactual (other arms) mediator assignment

Mediator	Control	Peer-Helper	Helpline	Peer-Helper + Helpline	Total
PH French-Arabic	78	66	91	63	298
PH Pashto A	53	47	37	64	201
PH Pashto B	54	51	43	43	191
PH Bengali A	24	44	36	45	149
PH Pashto C	41	24	36	24	125
PH Tamil-English	14	35	13	27	89
PH Bengali B	23	13	11	10	57
PH Arabic	1	5	3	5	14
PH Dari	2	1	4	4	11

Notes: Number of refugees per mediator and treatment arm. In the two peer arms the mediator is the one actually assigned; in the Control and Helpline arms it is the counterfactual mediator reconstructed by the assignment algorithm (language x gender priority list restricted to the mediators present on the treatment day, seeded random tie-breaks). Refugees whose actual mediator is not one of the nine listed mediators are assigned by the algorithm. Mediators are identified by the language they work in, never by name.

Table E11: Peer-assignment effect by (counterfactual) mediator

Mediator	Any post-treatment signal	Any hospital visit	N cell	N treated
PH French-Arabic	0.094**	0.027	207	129
	(0.046)	(0.042)		
PH Pashto A	-0.006	-0.023	164	111
	(0.039)	(0.037)		
PH Pashto B	0.010	0.012	148	94
	(0.047)	(0.047)		
PH Bengali A	0.067**	0.057**	113	89
	(0.026)	(0.024)		
PH Tamil-English	-0.097	-0.022	76	62
	(0.099)	(0.077)		
Other mediator	0.027	0.020	153	86
	(0.040)	(0.035)		
Observations	861	861		
p: equal effects across mediators	0.288	0.547		

Notes: Each cell of the top panel reports the effect of assignment to a peer arm (Peer-Helper or Peer-Helper + Helpline, versus Control) within one mediator cell, from a single OLS regression of the outcome on mediator-cell indicators and their interaction with the peer-assignment indicator, with randomisation strata fixed effects (site x nationality group x gender) and HC1 heteroskedasticity-robust standard errors in parentheses. The mediator cell is the actual mediator for peer-arm refugees and the counterfactual mediator reconstructed by the assignment algorithm for Control refugees, so treated and control refugees within a cell would face the same mediator; the five most frequent mediators are shown separately and the remainder grouped. The Helpline-only arm is excluded. The last footer row reports the p-value of the joint test that the peer effect is equal across mediator cells. * p<0.10, ** p<0.05, *** p<0.01.

RÉSUMÉ

Les programmes publics peuvent être formellement accessibles sans pour autant atteindre effectivement les populations auxquelles ils sont destinés. Cette thèse étudie en France trois dimensions de cette relation : l'engagement, l'allocation et l'intensité de l'intervention, et l'accès aux services.

Le premier chapitre étudie un programme proposant des ateliers de langue et d'intégration aux parents immigrés au sein des établissements scolaires de leurs enfants. Le programme n'améliore pas leur niveau de français, mais accroît leur participation aux activités scolaires et leur estime de soi. Leurs enfants progressent également de 0,25 à 0,30 écart-type aux évaluations de français, suggérant que l'engagement avec l'école peut constituer un canal efficace d'intégration.

Le deuxième chapitre porte sur les centres éducatifs fermés, créés comme une alternative à l'incarcération des mineurs. Parmi les mineurs pour lesquels un CEF a été recommandé, le placement se substitue principalement à des mesures de suivi en milieu ouvert plutôt qu'à la détention et augmente de 8 à 10 points de pourcentage la probabilité d'une détention dans les douze mois. Les résultats suggèrent que la révocation de la mesure judiciaire portant le placement constitue un mécanisme important.

Le troisième chapitre étudie l'accès des réfugiés aux soins de santé mentale. Dans une expérimentation randomisée menée auprès de 1 135 réfugiés, seule la combinaison d'un échange avec un pair-aidant et la mise à disposition d'une ligne téléphonique multilingue facilitant la prise de rendez-vous augmente le recours aux soins, de 6,2% à 13,7%, malgré un recours qui demeure faible.

Les résultats de cette thèse soulignent que la capacité des institutions à atteindre leurs publics dépend non seulement de ce que les programmes proposent, mais aussi de la manière dont sont organisés le contact, l'accompagnement et l'intervention.

MOTS CLÉS

Économie publique, Migration, Justice des mineurs, Santé

ABSTRACT

Public programmes can be formally accessible without effectively reaching the populations they are intended to serve. This thesis studies three dimensions of this relationship in France: engagement, allocation and intensity of intervention, and access to services.

The first chapter studies a programme offering language and integration workshops to immigrant parents inside their children's schools. The programme does not improve parents' French, but increases their participation in school activities and self-esteem. Their children also gain 0.25 to 0.30 standard deviations in French test scores, suggesting engagement with the school as an efficient channel of integration.

The second chapter studies closed educational centres, created as an alternative to juvenile incarceration. Among minors recommended for a CEF, placement mostly displaces community supervision rather than detention and raises the probability of detention within twelve months by 8 to 10 percentage points. The results suggest that revocation of the judicial order carrying the placement is an important channel.

The third chapter studies refugees' access to mental health care. In a randomised experiment among 1,135 refugees, only the combination of a peer-helper discussion and a multilingual helpline facilitating appointment booking increases engagement with care, from 6.2% to 13.7%, despite take-up remaining low.

Overall, the findings highlight that institutional reach depends not only on what programmes offer, but also on how contact, assistance and intervention are organised.

KEYWORDS

Public Economics, Migration, Juvenile Justice, Health

