ÉCOLE DES HAUTES ÉTUDES EN SCIENCES SOCIALES

DOCTORAL THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Three essays on intrahousehold dynamics and intimate partner violence (IPV) in West Africa

Publicly defended on February 25, 2022 by

SARAH DESCHÊNES

DISSERTATION COMMITTEE

External Reviewers

Alessandra VOENA	Professor, Stanford University
Rachel HEATH	Professor, University of Washington

Discussants

Tanguy BERNARD	Associate Professor, Université de Bordeaux

PhD Advisor

Sylvie LAMBERT	Senior Research Fellow, INRAE and Paris School of Economics
Marion LETURCQ	Research Fellow, INED

ÉCOLE DES HAUTES ÉTUDES EN SCIENCES SOCIALES

THÈSE POUR L'OBTENTION DU TITRE DE DOCTEUR EN SCIENCES ECONOMIQUES

Trois essais sur les dynamiques intrafamiliales et les violences conjugales en Afrique de l'Ouest

Thèse présentée et soutenue publiquement le 25 février 2022 par

SARAH DESCHÊNES

MEMBRES DU JURY

Directrices de Thèse

Sylvie LAMBERT	Directrice de Recherche INRAE et Paris School of Economics
Marion LETURCQ	Chargée de recherche INED

Rapporteurs

Alessandra VOENA	Professeure, Université de Stanford
Rachel HEATH	Professeure, Université de Washington

Suffragants

Tanguy BERNARD	Professeur Associé, Université de Bordeaux

A la mémoire de Florence et d'Angèle
Pour Régine Banda, Lucie et Maman sur les épaules de qui je me tiens debout aujourd'hui.
Pour Tristan

REMERCIEMENTS

C'est avec soulagement que je m'attelle enfin à la rédaction des remerciements! Cette thèse est le fruit de cinq années de travail, cinq années au cours desquelles j'ai rencontré des personnes brillantes et passionnées qui m'ont épaulée, inspirée et encouragée tout au long de cette thèse.

Je tiens d'abord à remercier Sylvie Lambert, qui, il y a six ans, a accepté d'encadrer mon mémoire de Master 2 à PSE, puis de devenir ma directrice de thèse. Nous nous sommes rencontrées peu avant la fin d'une année de césure pendant laquelle j'ai erré de sujet en sujet, passant d'un encadrant à un autre, avec le sentiment que mes difficultés à trouver le bon sujet et la bonne personne pour m'encadrer était peut-être un symptôme d'une inadéquation entre mon aspiration naissante à la recherche et ma capacité à devenir chercheuse. Sylvie, je te suis infinimment reconnaissante d'avoir fait le pari il y a cinq ans de m'accompagner dans ma formation à la recherche. Le fait que tu acceptes de devenir ma directrice de thèse alors que nous avions peu travaillé ensemble précédemment m'a insufflé de la confiance et a grandement contribué à affermir ma vocation. Au-delà de la bienveillance dont tu as fait preuve tout au long de notre collaboration, ta rigueur, ton intégrité, ton esprit critique et ton enthousiasme sincère sont des sources d'inspiration qui structurent ma vision de la recherche. Je tiens également à saluer la générosité avec laquelle tu as partagé avec moi ton experience ainsi que ton temps lors de passionnantes discussions. J'ai eu le plaisir d'être ta co-autrice, et j'espère avoir la chance de travailler de nouveau avec toi et de continuer à bénéficier de tes précieux conseils.

Je tiens également à remercier à titre posthume Véronique Hertrich (INED), qui a co-encadré ma thèse pendant une année avant que la maladie ne la contraigne à arrêter. La passion et l'exigence de Véronique sont des références vers lesquelles je tâche de tendre tout en concédant qu'il sera probablement difficile de les atteindre. Je remercie également Géraldine Duthé (INED) qui a pris le relais de Véronique pendant un an, puis Marion Leturcq (INED) qui m'a ensuite accompagné jusqu'à la fin de cette thèse. Marion,je te remercie pour ta disponibilité et pour le soutien que tu m'as apporté pendant la thèse. Je remercie également Liam Wren-Lewis qui a été un membre très impliqué de mon comité de thèse, et dont les précieux conseils, la gentillesse et la disponibilité ont été des atouts majeurs pour améliorer la qualité des papiers produits dans cette thèse. Je suis heureuse de savoir que nous allons travailler ensemble sur un projet de recherche et débuter une nouvelle phase de notre collaboration.

I would like to thank the two referees of my thesis, Alessandra Voena and Rachel Heath. I am very grateful for the care with which you read the papers, and the insightful comments you

made during the private defense. It undoubtedly improved the quality of the dissertation, and I hope that you will be satisfied with the current version of the manuscript.

Cette thèse a notamment été financée par l'Institut National des Etudes Démographiques (INED) qui m'a attribué un contrat doctoral de trois ans, puis un financement supplémentaire de six mois. Je remercie l'INED pour la confiance accordée. Je remercie également Denis Cogneau de m'avoir recommandé auprès de Léa Rouanet du Gender Innovation Lab (GIL) de la Banque Mondiale, ce qui m'a permis de financer mes deux dernières années de thèse.

Je tiens également à remercier tous les chercheurs que j'ai côtoyés à l'Ecole d'Economie de Paris et à l'INED au cours de ces cinq dernières anneées. Leurs conseils et leur soutien ont fortement contribué à mon apprentissage de chercheuse. Je pense en particulier à Denis Cogneau, Karen Macours, Luc Behaghel, Oliver Vanden Eynde, Kenneth Houngbedji (inoubliable TA du cours de Measurement en Master PPD), Akiko Suwa-Eisenmann à PSE; à Valérie Golaz, Géraldine Duthé, Marion Leturcq et Dominique Meurs à l'INED. Je souhaite particulièrement remercier Denis, dont la dévotion pour la recherche est une inspiration depuis que j'ai pu bénéficier de son enseignement en Master PPD. Je remercie également Léa Rouanet du Gender Innovation Lab (GIL) de la Banque Mondiale et Rachel Cassidy pour leur management efficace et bienveillant, et leur soutien pendant ces dernières années de thèse.

Ma thèse doit énormément à mes co-autrices, en particulier Rozenn Hotte. Rozenn, j'admire ta passion pour la recherche, ta détermination et ton intelligence. Je te remercie de m'avoir proposé de travailler avec toi sur ce projet sur la scolarisation primaire au Bénin. En plus des longues et passionnantes discussions avec toi, nous avons tissé des liens d'amitié qui m'ont permis de tenir le coup dans les moments difficiles et qui ont aussi décuplé les moments de joie. J'ai également eu le plaisir de co-écrire avec Sylvie et Christelle Dumas que je remercie de nouveau pour leur générosité lors de notre collaboration. Je souhaite remercier Victor Pouliquen, qui m'a permis d'arrimer mon expérience de liste à l'un de ses projets de recherche au Burkina Faso. Le contenu de cette thèse aurait été radicalement différent sans son aide précieuse. Victor, tu ne t'es pas limité à m'autoriser à rejoindre ton projet, tu as aussi été très généreux dans ton aide et tes conseils pour m'aider à aborder mon travail de terrain. Je remercie également Carole Treibich pour ses nombreux conseils sur la mise en place de l'expérience de liste, et Amber Peterman pour avoir eu la générosité de partager des éléments de son propre protocole de recherche avec moi. Cette thèse doit également à la qualité et l'implication de nombreuses enquêtrices au Burkina Faso, en particulier celles qui ont participé au pilote de l'expérience de liste; Estelle Ouleane, Foussena Sawadogo, Bintou Ouedraogo, Marcelle Kiema et Diane Kaboré, avec une mention particulière à Diane, qui a été présente des premiers entretiens qualitatifs à la collecte de données dans l'ensemble du Burkina. Je remercie également Françoise Ki-Zerbo dont l'hospitalité et la gentillesse ont rendu mon arrivée à Ouagadougou si plaisante.

Merci à mes amis à PSE et à l'INED d'avoir rendu le quotidien de la thèse si plaisant ! Un merci tout particulier à Paola, Rozenn, Yasmine, Lisa et Juliette pour la super ambiance,les encouragements et la bienveillance qui règne dans le bureau R6-01. Paola, merci beaucoup pour ta patience et tes excellents conseils. Juliette, je ne sais même pas par où commencer pour te

remercier. Tu as été ma co-bureau pendant ces cinq années de thèse, nous avons fait ce bout de chemin ensemble et rien n'aurait été pareil sans ta présence. J'ai toujours pu compter sur toi dans les bons moments mais aussi pendant les périodes difficiles. Nous avons probablement bu des centaines d'hectolitres de thé et de tisane pendant cette thèse! Merci pour les fous rires, les relectures de mails et de papiers, les discussions passionnantes que nous avons eu et qui ont souvent mené à des prises de conscience importantes, tant scientifiques que personnelles. Merci pour tes conseils, toujours pertinents, tes partages de lectures, les vidéos de yoga pour décompresser quand l'angoisse de la thèse prenait le dessus. Merci pour ton soutien indéfectible, ta générosité, ta bienveillance et ton écoute qui ont illuminé mes années de thèse. Un grand merci à la nouvelle génération des occupants du bureau R6-01 qui ont repris le flambeau de l'efficacité bienveillante avec brio: Duncan, Zhexun, Andréa et Charlotte, le covid-19 nous a privé d'une année à travailler et rire ensemble mais j'espère vous retrouver bientôt. Bien d'autres doctorants ont enrichi mon quotidien, et je les en remercie: Alessandro, Giulio, Emmanuele, Marion, Paul, Justine, Eric, Louise, Mélanie, Irène et Ismaël. Un clin d'oeil tout particulier à Juni, qui a été un rayon de soleil et un puit de motivation, en particulier dans les dernières années de thèse.

J'ai également été très heureuse d'avoir la chance de travailler à l'INED; cela m'a permis d'être exposée à une approche pluri-disciplinaire à laquelle ma thèse doit beaucoup, de découvrir des recherches passionnantes, et surtout de rencontrer des personnes extraordinaires. Je tenais tout particulièrement à remercier Benjamin avec qui j'ai beaucoup ri. Merci aux autres doctorants pour leur bienveillance et leur implication dans l'amélioration des conditions de travail des doctorants: Julie, Maxime, Elisa, Noémie, Martha, Matthias et Amélie(s).

Je souhaite également remercier le personnel administratif de PSE, en particulier Radja Aroquiaradja, Stéphane Brice, Marie Font, l'unique Véronique Guillotin, Béatrice Havet, Eliane Madeleine, José Sastre, et Pauline Marmin, et Léa Bourset, et Adriana Banta à l'INED.

Enfin, je remercie ma famille et mes amis, qui m'ont apporté un soutien sans faille, même lorsque la physionomie de mes journées de travail paraissait bien mystérieuse à certains. Un grand merci à Constance et Thomas, et maintenant Raphaël, pour les réconfortantes soirées bucket ! Merci à Sophie, Juliette, Marie, Elise, Safaa, Misung pour les apéros, dîners et brunchs! Un grand merci à Clémence pour sa bienveillance, sa franchise et ses encouragements depuis toutes ces années.

Merci à Michel de m'avoir montré son soutien dès que j'ai manifesté l'envie de faire de la recherche. Merci à Caroline pour son soutien, son écoute attentive et ses bons petits plats, et merci également à Etienne. Merci à mes parents pour leur confiance et leur soutien, même quand mon choix de me réorienter après l'école de commerce les a désarçonnés. Lucie, si je dois énumérer les raisons pour lesquelles je te suis reconnaissante, je suis partie pour rédiger une encyclopédie. Je distingue très lucidement ton influence sur la chercheuse que je suis devenue; dans l'importance que j'accorde au fait de s'investir pleinement dans ce que je fais, dans ma soif d'apprendre et dans l'attention portée aux conséquences de mes actions sur le monde qui m'entoure. Merci également à Samuel.

Enfin, merci à Tristan, avec qui la vie est une belle aventure, de Jouy-en-Josas à Cotonou, en passant par Paris et Chicago. Merci de m'avoir fait tant rire pendant toutes ces années, de m'avoir rassurée dans les moments de stress, parfois même de détresse, et de m'avoir aussi rappelé que la vie ne se limite pas à mon travail de recherche. Je te dois bien plus que je ne peux l'écrire ici.

ABSTRACT

Les familles africaines sont traversées par des dynamiques complexes et sont lieu de rapport de domination de la femme par l'homme, des jeunes par les aînés. Comprendre plus finement le fonctionnement des familles africaines permet d'identifier les sources de fragilité et de force de leurs membres dans un contexte où ces derniers comptent sur la structure familiale pour leur fournir un filet de sécurité informel. Les violences domestiques dont les principales victimes sont les femmes, font partie du quotidien des familles africaines et sont un reflet de la domination masculine dans des sociétés où les hommes demeurent les piliers du lignage. Dans cette thèse d'économie, j'étudie dans quelle mesure le risque pour une femme d'être victime de violences domestiques est atténué par son niveau d'éducation et par la présence d'alliés renforçant son statut dans son réseau familial.

Dans le premier chapitre de la thèse, j'examine l'effet du capital humain acquis à l'école primaire sur des indicateurs de qualité de la vie conjugale des femmes au Bénin. Dans ce travail co-écrit avec Rozenn Hotte, j'exploite la forte augmentation des constructions d'écoles dans les années 1990 pour évaluer l'impact causal d'une augmentation de l'offre scolaire primaire sur la fréquentation de l'école primaire, l'âge au mariage, l'âge à la première naissance, ainsi que la tolérance et l'expérience des violences conjugales. Nous utilisons les variations géographiques et historiques quasi-expérimentales du nombre d'écoles construites, et nous trouvons que le programme de construction d'écoles a augmenté la probabilité de fréquenter l'école primaire dans les zones rurales, a augmenté l'âge au mariage et a diminué la probabilité que les femmes tolèrent la violence conjugale. Les femmes plus éduquées sont également moins à risque d'être exposées à la violence de leur conjoint. Les effets rapportés dans l'étude passent par l'éducation des femmes, ils ne résultent pas d'un changement dans l'éducation de leur mari.

Dans le deuxième chapitre de la thèse, j'étudie si les femmes mariées dont le premier enfant est un fils sont moins susceptibles de tolérer et de subir des violences conjugales au Burkina Faso. Le mécanisme clé est l'existence d'une préférence pour les garçons qui peut renforcer le

statut des mères de garçons au sein de la famille. En utilisant les EDS de 2003 et 2010, j'exploite l'exogénéité du sexe du premier enfant et, le cas échéant, le fait que les femmes sont dans une union polygame pour utiliser comme stratégie d'identification un effet fixe du mari. L'analyse ne permet de pas de fournir une réponse tranchée à la question de recherche. L'impact du sexe du premier enfant sur sa mère varie en fonction du type d'union (monogame ou polygame), et de son rang parmi les coépouses pour les femmes mariées à un partenaire polygame. Je trouve notamment des preuves descriptives que les premières épouses dont l'aîné est un fils bénéficient d'une plus grande autonomie dans la prise de décision, et que, pour les secondes épouses, le fait d'avoir un fils est corrélé avec une moindre tolérance à des violences conjugales et avec un accès accru aux ressources, mesuré par la probabilité d'être anemiée. Les données sont également compatibles avec l'hypothèse que plus une femme reste longtemps sans avoir de fils, plus elle est exposée à la violence. Pour les femmes monogames, le fait d'avoir un fils peut conduire à un comportement violent de la part du mari. La complexité des résultats descriptifs de l'étude questionne la qualité des outils de mesure de la tolérance des violences conjugales, de la participation aux décisions et de l'expérience de la violence domestique utilisées dans les grandes enquêtes auprès des ménages.

Dans le troisième chapitre de ma thèse, j'étudie les chiffres cachés de la violence domestique dans les zones rurales du Burkina Faso et je mets en avant les caractéristiques des femmes qui subissent des violences conjugales mais qui ne le révèlent pas dans les grandes enquêtes auprès des ménages. Je compare la prévalence des différents types de violence conjugale obtenue lorsque les femmes sont interrogées directement sur leur expérience de la violence (par exemple, "avez-vous déjà été giflée par votre partenaire") à une mesure indirecte des violences obtenue par une expérience de liste[1]. Je mesure la prévalence de plusieurs types de violence conjugale (violence physique moins grave, violence physique grave et viol conjugal) avec les deux méthodes et je constate que les mesures directes des violences conjugales sous-estiment les formes les plus intenses de violence de 7 à 9 points de pourcentage. Je ne trouve pas de biais pour la violence physique moins sévère. Je constate également que le fait d'avoir une fille aînée présente une corrélation différente avec les violences conjugales selon la mesure de la violence utilisée.

[1]Le principe de l'expérience de liste est le suivant. On présente à un groupe d'individus (groupe A) une liste de trois éléments non sensibles, par exemple, et on leur demande combien d'éléments ils approuvent au total. Ensuite, on présente à un autre groupe (groupe B) la même liste de trois éléments non sensibles à laquelle on ajoute l'élément sensible, qui représente l'opinion ou le comportement sensible que le chercheur veut mesurer. Grâce à la répartition aléatoire des répondants dans les groupes A ou B, la différence entre le nombre moyen d'éléments sur lesquels les participants du groupe B sont d'accord et le nombre moyen d'éléments sur lesquels les participants du groupe A sont d'accord devrait révéler la véritable prévalence de l'opinion ou du comportement sensible.

**** *English version*

African families are subject to complex dynamics and are the site of power relationships that give men privileges over women, and gives elders privileges over young people. A more detailed understanding of the functioning of African families makes it possible to identify the sources of fragility and strength of their members in a context where the latter rely on the family structure to provide them with an informal safety net. Domestic violence, whose main victims are women, is part of the daily life of African families and is a reflection of male domination in societies where men remain the pillars of the lineage. In this economics thesis, I will study the extent to which a woman's risk of becoming a victim of domestic violence is mitigated by her level of education and by the presence of status-enhancing allies in her family network.

In the first chapter of the dissertation, I examine the effect of human capital acquired in primary school on women's marital outcomes in Benin. In this joint work,[2] I exploit a sharp increase in school constructions in the 1990s to assess the causal impact of an increase in primary school supply on primary school attendance, age at marriage, age at first birth and the tolerance and experience of IPV. Using the quasi-experimental geographical and historical variations in the number of schools built, we find that the school building program increased the probability of attending primary school in rural areas, increased age at marriage and decreased the probability of women being tolerant of wife beating, and decreased their experience of IPV as well. These effects are going through women's own education and do not result from a change in their husband's education.

In the second chapter of the dissertation, I study whether married women whose first child is a son are less likely to tolerate and experience IPV in Burkina Faso. The key mechanism is the existence of a preference for sons that may strengthen the status of the mothers of boys within the family. Using the 2003 and 2010 DHS, I exploit the exogeneity of the sex of the first child and, when applicable, the fact that women are in a polygamous union to apply a husband-fixed effect strategy. I am unable to provide a clear-cut answer to the research question. The impact of the sex composition of children on the mother's tolerance and experience of IPV varies with the type of union (monogamous or polygamous), and with the rank of the spouse among women married to a polygamous partner. I find descriptive evidence that first wives whose eldest child is a son benefit from more autonomy in decision-making, and that, for second wives, having a son is correlated with less tolerance of IPV and with an increased

[2] with Rozenn Hotte, postdoctoral researcher at the THEMA-CY Cergy Paris University

access to resources proxied by the probability to be anemic. I also find suggestive evidence that the longer a wife remains without a son, the more domestic violence she might experience, but that, for monogamous women, once a son is born, it might also lead to a violent behavior from the husband. The complexity of the descriptive results the study yields calls into question the quality of the measure of tolerance of IPV, participation in decision-making, and of experience of IPV used in large household surveys.

In the third chapter of my PhD, I study the hidden figures of domestic violence in rural Burkina Faso and uncover the characteristics of women who experience IPV but do not reveal it in large household surveys. I compare the prevalence of IPV obtained when women are asked directly about their experience of IPV (for instance, "have you ever been slapped by your partner") to an indirect measure of IPV obtained with a list experiment.[3] I measure the prevalence of several types of IPV (less severe physical violence, severe physical violence and marital rape) with both methods and find that the direct measures of IPV underestimate the most intense forms of violence by 7 to 9 percentage points. I find no bias for less severe physical violence. I also find that having a first-born daughter and being exposed to a radio campaign meant to increase modern contraception take-up correlate differently with IPV according to the measure of IPV used. To the best of my knowledge, this paper is among the first to study the importance of hidden domestic violence in a Western African country.

Keywords: family, gender, intimate partner violence (IPV), son preference, children, marriage, education, sub-Saharan Africa.
JEL Classification: O12, J12, J16, I25, I32, D19, C99, O55.

[3]The idea behind a list experiment is as follows. One group of individuals (group A) is presented with a list of three non-sensitive items, for instance, and is asked how many items they agree with in total. Then, another group (group B) is presented the same list of three innocuous items and a sensitive item, which is the sensitive opinion or behavior the researcher wants to measure. By virtue of the randomization, the difference in the mean number of items agreed upon in group B minus in group A should reveal the true prevalence of the sensitive opinion or behavior.

Contents

List of Figures

List of Tables

General Introduction

The complexity of intra-family relations in Africa is the subject of growing interest in economic literature. Collective models are among the tools developed to understand the complexity of family relations. They are, however, built for nuclear households, and they most often make the hypothesis of cooperation between the members of a household. However, cooperation between the members of a family is not obvious, especially when the family composition is more complicated than that of the nuclear family. A more detailed understanding of how the family works is critical to identify the sources of fragility and strength of their members in a context where they rely on the family structure to provide then with a safety net. The power struggles within families can affect household members differently, particularly according to their gender or their age. The patriarchal organization of many African societies leads to situations where women tend to be in a position to be dominated by their husbands or their husbands' families at different stages of their life cycle. This imbalance is reflected in the existence of domestic violence against women. Domestic violence is analyzed in the dissertation as a symptom of the institutionalization of male domination over women. Domestic violence is part of the daily life of African families. According to the World Health Organization (WHO), in 2018, 33% of ever-partnered women had experienced IPV in Sub-Saharan Africa.[4] In Burkina Faso, according to Demographic and Health Surveys (DHS) 2010, nearly one-third of women believe that it is justified for a husband to beat his wife if she goes out without telling him. This proportion reaches 41% in Senegal (DHS 2014) and 55% in Mali (DHS 2012-2013). In addition, 45% of Burkinabe women surveyed said they could not refuse sex from their spouse or partner if They do not wish to have sex.

The objective of the dissertation is to advanced the understanding of the sources of fragility

[4]Compared to 33%with South-East Asia, 20%in the Western Pacific, 22% in high-income countries and Europe and 25% in the Americas.

and strength of women in their household. The literature shows that family composition or access to resources that are supposed to contribute to women's empowerment have ambiguous effects on women's empowerment and on domestic violence. What may appear to be favourable to women's empowerment can raise other difficulties that may counteract the expected beneficial effects. For instance, Bobonis *et al.* (2013) show that in Mexico, women who receive cash transfers through the Oportunidades program are less likely to be physically abused by their spouses, but are more likely to be verbally abused by them, and to separate from their partner. In another context, Anderson and Genicot (2015) show that an increase in property rights for women in India is associated with more suicides for both women and men, as it increases the conflictuality of unions. The increase in property rights is also associated with a higher incidence of domestic violence against women. An empirical approach is therefore needed to determine whether the dominant effect is positive or negative, and in which contexts. Once this is understood, it is then possible to integrate this dimension into the design of emancipatory policies or, at the very least, to identify women at risk of IPV and to set up educational actions aimed at men to protect vulnerable women. This work is original because it is at the intersection of two areas of economic literature with strong potential for development: the theme of resource sharing within the household and more generally of relations between individuals within the household, a theme over which is superimposed the question of the determinants of the negotiating power and status of individuals, which may have an effect on domestic violence.

In the first chapter of the dissertation, I explore whether the conditions in which women start their marital life impacts the quality of their marital outcomes, including their tolerance and experience of IPV in Benin. More precisely, I study whether receiving more primary education and being more educated upon marriage makes women better off. Even though the benefits of education for women are undisputed, the link with tolerance and experience of IPV is not as straightforward as it may seem. In a context where education is of low quality, attending school may not necessarily lead to improvement in women's acess or earnings on the labor market. In addition, if education is a factor of empowerment, being an empowered spouse may result in women acting in a gender-incongruent way, which may expose them to backlash. Conversely, as evidenced by Bursztyn *et al.* (2017), women may further internalize the gender norms that constrain their behavior, and, in a context where women are expected to show deference to their husband (Falen (2011)), downplay their support to gender progressive norms in order to abide by the attitude expected from a wife in their environment. In this joint work with Rozenn Hotte, we build on the existing literature on education and women's

marital outcomes and explore further the channels through which primary education impacts women's tolerance and experience of IPV, including potential changes in the quality of a key family player in the occurence of IPV: the husband.

After studying how education may shape the way women start and experience their marital life, chapter 2 of the dissertation studies whether in Burkina Faso the gender composition of children impacts the tolerance of IPV of the mother, her autonomy in the decision-making in the household, her health, and her experience of IPV. Even though the preference for sons in West Africa is more subtle than in India or China, in contexts of patrilinearity, filial relations are a structuring axis of the power relations in the family. Children represent a capital that allows a woman to assert herself, especially in contexts where members of a household compete to access limited resources, like when the head of household is polygamous. This aspect of family relations has recently been taken into account in work that examines the relationship between family status and violence, as well as in the developing economic literature on household relations and resource sharing. Yount (2005) studies the role of family composition on domestic violence in the Minya governorate of Egypt. In particular, she analyzes the impact of a woman's number of sons on her acceptance of domestic violence and on the violence she actually experiences, and observes the role of cohabitation among family members. Her work concludes that married women in a situation of dependence on their husbands (dependence being defined as having a lower level of education than the husband and having several boys) are more vulnerable. Women who are isolated from their maternal family also exhibit a greater likelihood of being a victim of domestic violence. Yount (2005) sees the number of boys as a burden and a factor of fragility for the wife. However, we can take another look at the effect of male births and think that in societies with a preference for boys, being the mother of boys would reinforce the status of the woman (Rossi and Rouanet (2015) highlight such a preference in Egypt for example). For example, Lambert and Rossi (2016) show that in Senegal, women whose husbands have children from a previous union and are at risk of becoming widows, modify their fertility behavior by reducing the birth gap until they have a son who provides them with a form of insurance if their husbands should die.

In the third chapter of the dissertation, I take a step back to reflect on the tools used to measure IPV. In Economics in particular, IPV is measured using direct questions to women about precise acts of violence they may have experienced. The experience of IPV is a very sensitive phenomenon, however, which makes it difficult to measure acurately (Tourangeau and Yan (2007)). In the context of West Africa in particular, women are expected to appear obedient

and loyal to their husband, to salvage his image regardless of what is happening behind closed doors (Falen (2011)). Researchers therefore expect that the measuring IPV by directly asking women to report their experience is likely to lead to underreported measures of IPV. Worst, direct questions on IPV may bias in a nonrandom way the prevalence of IPV, making it difficult to draw lessons from policies meant to improve women's outcomes. Though the literature acknowledges that measures of IPV may be biased, the attempts to find alternative measures of IPV remain limited. In the third chapter of the dissertation, I test an alternative way of measuring IPV: the list experiment (LE). It is a technique that measures sensitive behaviors or attitude in an indirect way, and that conceals the answer of the respondent, providing them with a certain level of plausible deniability. It is meant to eliminate the social desirability bias that may arise from inquiring about a sensitive topic (Droitcour *et al.* (2004)). Using the LE allows me to understand the magnitude of IPV that is missed with the direct measure of violence, and the characteristics of the women that we miss in the large household surveys widely used in the literature on IPV.

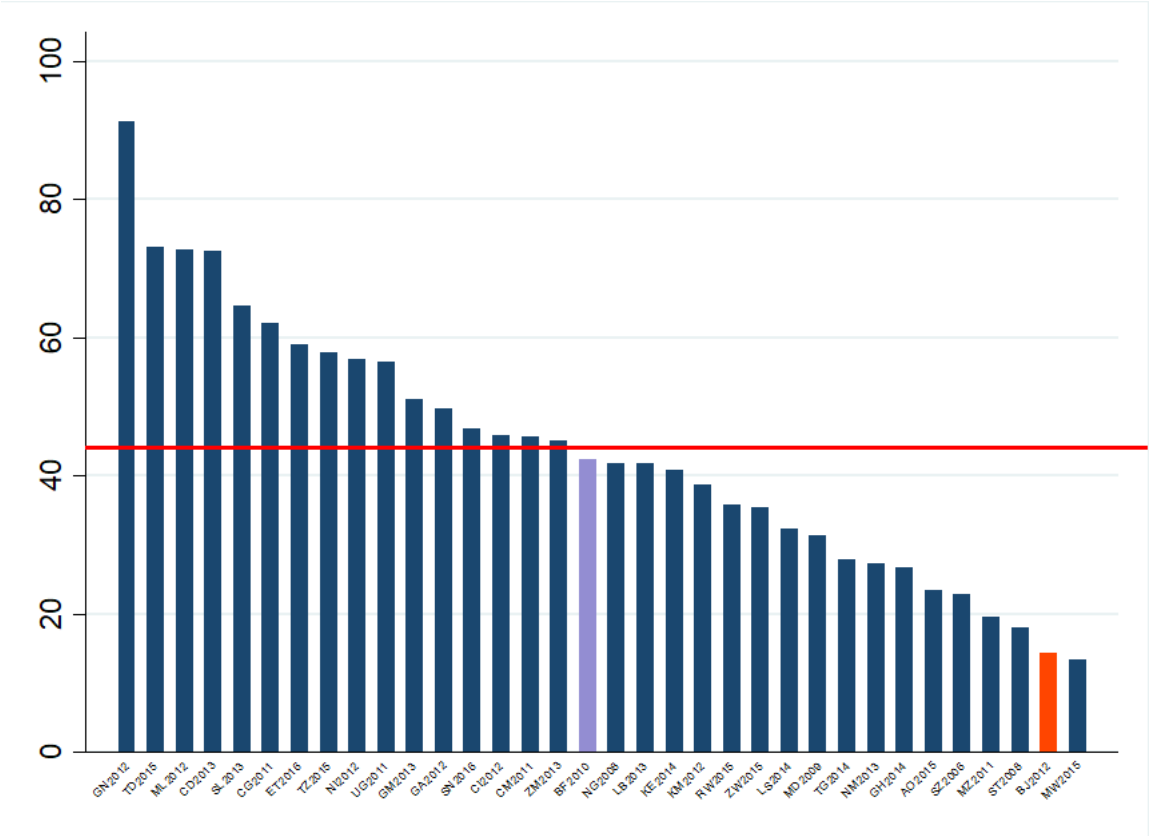
In this general introduction, I start by providing a brief overview of the levels of tolerance and experience of IPV in sub-Saharan countries using data from the Demographic and Health Surveys (DHS). Then, I discuss the conceptual frameworks used by economists to comprehend resources sharing in the household and the part that domestic violence may play in securing resources. I then present the data used in this PhD dissertation, and discuss some personal experience from the field. In order to illustrate how my work on the measure of IPV is part of a tradition of debates around the measure of domestic violence, I provide a selective synthesis of the arguments of the debates on the tools used to measure IPV, especially the direct measure of IPV. It aims at showing that even a widely accepted measure of IPV among economists remains the object of intense debates in other disciplines. Finally, I outline the structure of the dissertation.

Levels of tolerance and experience of IPV

Based on the DHS data collected in 33 countries of sub-Saharan Africa between 2008 and 2015, on average 43.66% of women found wife-beating justified for neglecting the children, arguing with the husband or going out without telling the partner. As shown by figure 0.1, the average —indicated by the red line —conceals high variation across country, with a rate as high 91% in Guinea in 2012 to a rate of 9% in 2015 in Malawi. The rectangle in purple and in orange indicate the tolerance of IPV for the described motives in Burkina Faso in 2010, and in Benin in 2012. With the rate of 42%, Burkina Faso in 2010 is just below the mean, whereas a rate of

14% places Benin among the countries with the lowest tolerance of IPV. Figure 0.2 shows that there is also important variations across countries when it comes to justifying wife-beating for refusing sex to one's partner. Within West Africa, there are important variations as well, from countries like Benin exhibiting a rate of 7% in 2012 to countries like Mali in 2012 or Chad in 2015 exhibiting among the highest rate on the continent (60% and 44%). In Burkina Faso in 2010, 20% of women believed that a partner was entitled to beat his wife if she refuses to have sex with him.

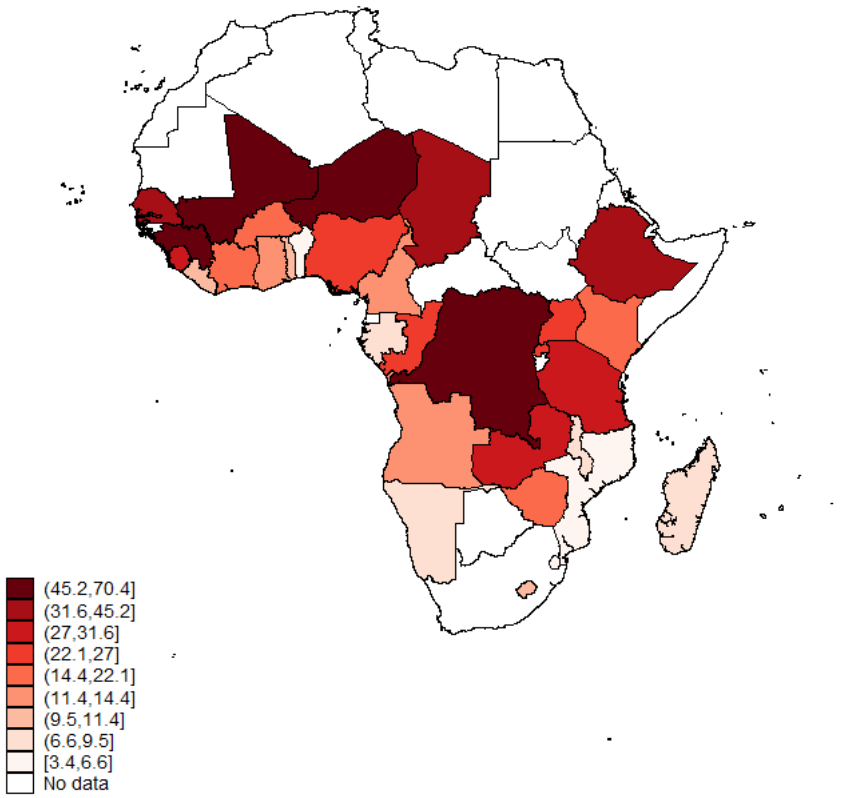
Figure 0.1: Tolerance of IPV for common offense in sub-Saharan Africa



Note: The figure presents share of women who agree that wife beating is justified if a wife goes out without telling her partner, neglects the children, or argues with the partner. Sample: ever-partnered women aged 15-49 years old. **Source**: DHS, Angola 2015, Burkina Faso 2010, Benin 2012, DRC 2013, Congo 2011, Côte d'Ivoire 2012, Cameroon 2011, Ethiopia 2016, Gabon 2012, Ghana 2014, Gambia 2013, Guinea 2012, Kenya 2014, Comoros 2012, Liberia 2013, Madagascar 2009, Mali 2012, Malawi 2015, Mozambique 2011, Nigeria 2008, Niger 2012, Namibia 2013, Rwanda 2015, Sierra Leone 2013, Senegal 2016, Sao Tome and Principe 2008, Swaziland 2006, Chad 2015, Togo 2014, Tanzania 2015, Uganda 2011, Zambia 2013, Zimbabwe 2015.

On average, in the 26 countries for which the data is available, 20% of women experienced a form of emotional violence, 28% of women have been slapped by their partner, thrown at an object that could have hurt them, or have been shaken or pushed by their partner (classified as less severe forms of IPV), 12% have already been punched, kicked, dragged to the grond

Figure 0.2: Tolerance of IPV for refusing sex in sub-Saharan Africa

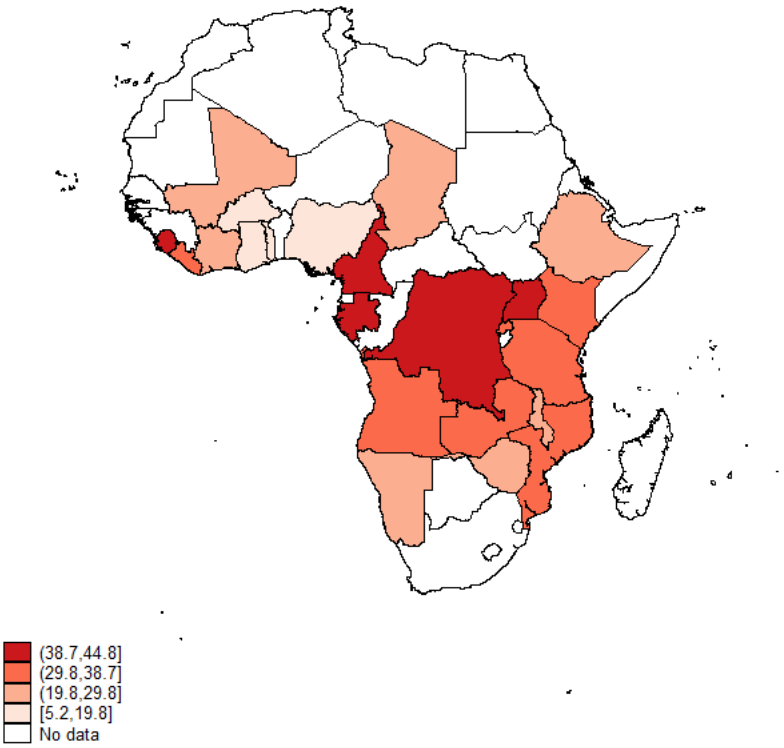


Note: The figure presents share of women who agree that wife beating is justified if a wife goes out without telling her partner, neglects the children, or argues with the partner. Sample: ever-partnered women aged 15-49 years old. **Source**: DHS, Angola 2015, Burkina Faso 2010, Benin 2012, DRC 2013, Congo 2011, Côte d'Ivoire 2012, Cameroon 2011, Ethiopia 2016, Gabon 2012, Ghana 2014, Gambia 2013, Guinea 2012, Kenya 2014, Comoros 2012, Liberia 2013, Madagascar 2009, Mali 2012, Malawi 2015, Mozambique 2011, Nigeria 2008, Niger 2012, Namibia 2013, Rwanda 2015, Sierra Leone 2013, Senegal 2016, Sao Tome and Principe 2008, Swaziland 2006, Chad 2015, Togo 2014, Tanzania 2015, Uganda 2011, Zambia 2013, Zimbabwe 2015.

of threatened with a weapon by their partner (classified as severe forms of IPV), and 10% of women experienced marital rape. As for tolerance of IPV, figure 0.3 and figure 0.4 shows that the averages conceal significant differences across countries, and regions of the continent.[5] In Burkina Faso in 2012, 9% of women experienced emotional violence, 9.5% of women experienced less severe physical IPV, 2% experienced severe IPV, and 1.5% declared they experienced marital rape.

[5] As the module on IPV is optional, the countries that choose to implement it may be selected.

Figure 0.3: Less severe physical violence in sub-Saharan Africa



Note: The figure presents share of women whose partner has already slapped them or thrown at them an object that could have hurt them, or pushed them or shook them. Sample: ever-partnered women aged 15-49 years old. **Source**: DHS. Angola 2015, Burkina Faso 2010, DRC 2013, Côte d'Ivoire 2012, Cameroon 2011, Ethiopia 2016, Gabon 2012, Ghana 2008, Gambia 2013, Kenya 2014, Comoros 2012, Liberia 2007, Mali 2012, Malawi 2015, Mozambique 2011, Nigeria 2008, Namibia 2013, Rwanda 2015, Sierra Leone 2013, Sao Tome and Principe 2008, Swaziland 2006, Chad 2015, Togo 2014, Tanzania 2015, Zambia 2013, Zimbabwe 2015.

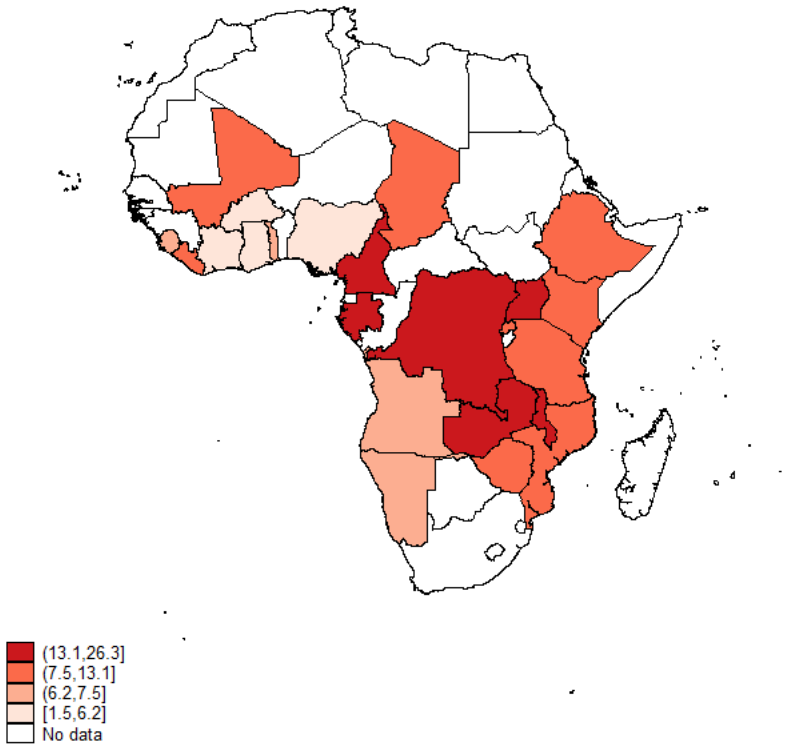
Conceptual Framework

Household models in Economics and Domestic violence

Before diving in the details of the section, it is important to note that the structure of the family in West Africa is complex. The concept of the nuclear family remains quite remote from the set up of West African families, which usually consist of a network of multiple households, related by blood or marriage but that do not necessarily live together in the same dwelling. Some of the models presented below were first developed to account for the inner workings of nuclear families, in the context of the US or Europe, not for the non-nuclear families of West Africa.

In a joint work with Christelle Dumas and Sylvie Lambert (Deschênes *et al.* (2020)), I reviewed the literature that offers insights into the intra-household decision-making process, the

Figure 0.4: Marital rape in sub-Saharan Africa



Note: The figure presents share of women whose partner has already forced them to have sexual intercourse when they did not wanted to do or to perform sexual acts they did not want to perform. Sample: ever-partnered women aged 15-49 years old. **Source**: DHS. Angola 2015, Burkina Faso 2010, DRC 2013, Côte d'Ivoire 2012, Cameroon 2011, Ethiopia 2016, Gabon 2012, Ghana 2008, Gambia 2013, Kenya 2014, Comoros 2012, Liberia 2007, Mali 2012, Malawi 2015, Mozambique 2011, Nigeria 2008, Namibia 2013, Rwanda 2015, Sierra Leone 2013, Sao Tome and Principe 2008, Swaziland 2006, Chad 2015, Togo 2014, Tanzania 2015, Zambia 2013, Zimbabwe 2015.

strategies put in place by individuals to secure their access to private resources, and discussed the place of IPV in these strategies.

The models used by economists to study the family as an economic unit have mainly relied on the vision of family members as cooperative and altruistic individual who share resources in a way that satisfies a set of family preferences. The most common representation of the household is the unitary model, which assumes that the household acts as a single individual. However, the unitary model has attracted many criticisms, in particular for failing to take into account the gender power balance within the household (Folbre, 1986).[6] Among cooperative models, an alternative representation of the household is provided by the collective

[6]Furthermore, the hypothesis of full income pooling that is required by the unitary model has been proven to be inadequate in both developed (Browning *et al.*, 1994) and developing countries ((Duflo, 2003; Hoddinott and Haddad, 1995; Thomas, 1994)).

model that represents the household as a set of individuals who make decisions through a bargaining process. The collective model relies on the hypothesis that households' decisions are Pareto optimal.[7] This model sees decision making as a two-step process, whereby public good consumption is decided and the remaining resources are shared among decision makers (in general, the two members of a couple) for their private consumption. The sharing depends on each household member's bargaining power, which is affected by the distribution factors that determine each member's outside options.[8] [9]

Though key to reflect on resource sharing in the household, cooperative models leave little room to include a discussion on domestic violence. Bargaining models (Manser and Brown (1980) and McElroy and Horney (1981) for instance), however, lay interesting foundations for economic models to include IPV, as they allow unions to breakdown as the result of the bargaining process. In this approach, the household is a framework in which spouses bargain about resource sharing. If spouses disagree, each spouse can stop bargaining. The solution of the model provides each spouse with a level of utility at least as high as the one she or he would get outside the marriage. The level of utility that an individual gets outside the marriage is the threat point in the bargaining process, and the higher, the more credible the threat. The level of the threat point can be influenced by many things including the wife's income, the laws that determine divorce settlements, or the probability to find a new partner.

Some non-cooperative models of the household have been elaborated to account for domestic violence, but a general theoretical framework is still lacking. In an original approach, Bloch and Rao (2002) uses ethnographic information on dowry-related violence in India to design a model of non-cooperative household bargaining with asymmetric information, wherein husbands use domestic violence as a signal of their dissatisfaction with the marriage. Husbands have an incentive to use violence against their wife to extract additional resources from her family. When the wife's family is richer (and the husband is less satisfied with the union), the

[7] A situation is said to be Pareto optimal or Pareto efficient if no change in the allocation of resources could lead to improved satisfaction for all parties. "Pareto efficiency" is considered as a minimal notion of efficiency that does not necessarily result in a socially desirable distribution of resources: it makes no statement about equality, or the overall well-being. In a couple, if the allocation of resources is not Pareto optimal, it means that there would be a way to improve the satisfaction of one person without taking anything away from the other one. This would suggest that some resources are wasted rather than being distributed within the couple. A reallocation of resources that reduces this "waste" brings the couple closer to the optimal situation, even if it is not sufficient to attain Pareto efficiency.

[8] Sen (1990) proposed an early discussion of models of cooperative conflict in the household.

[9] A number of papers that directly test the restrictions imposed by the model in developing countries suggests that this model might not typically hold. In fact, the Pareto efficiency hypothesis is not always found to be valid in such contexts, which implies that households are wasting resources (Udry (1996), De Mel *et al.* (2009), Duflo and Udry (2004), Dercon and Krishnan (2000) and Robinson (2012)). Baland and Ziparo (2018) present a clear overview of the empirical regularities observed in developing countries that suggest strategic decision making.

incentive to exert physical violence is stronger. Following the approach of Bloch and Rao (2002), Anderson and Genicot (2015) develop a model of intra-household bargaining with asymmetric information, where conflict is considered to be an intrinsic and likely part of bargaining. The model predicts that when wives contribute a greater share of the household wealth, they expect and obtain a higher share of the consumption. However, as a consequence, there may be more conflict in the household. An empirical test confirms the predictions of the model. An improvement in women's access to property induced by a change in the law is associated with an increase in both female and male suicide rates, potentially as a consequence of more intra-household conflict.[10]

Finally, inspiration can be found in the literature on identity economics to renew theoretical approaches of IPV. Akerlof and Kranton (2000) discusses a theoretical framework in which both partners derive utility from their self-image when it is consistent with a reference, for instance, traditional gender roles. The consistenty with traditional gender roles directly impacts the utility that they derive from marriage and, may shape the male partner's incentive to use physical violence against his wife.

Conceptualizing the motives of domestic violence

Both the theoretical and the empirical literature demonstrate that men develop non-cooperative strategies to prevent women from accessing resources within the household (see Deschênes *et al.* (2020) for a selective review). In the household, the range of action at men's disposal goes from simple obstruction to, at other end of the spectrum, exerting violence against their partner, whether physical or emotional, to thwart women's attempts to access resources. Although I do not assert that women necessarily experience violence because of disagreements that arise from intra-household bargaining, male-partner violence is nevertheless a behaviour that can occur during this process. There are two main motives that have been formalized by the sociological literature to attempt to explain why men resort to physical coercion: instrumental violence and expressive violence. Instrumental violence assumes that husbands use violence to control their partner's behaviour and access to resources in the household. In cases of expressive violence, the partner uses violence to vent frustration or stress and directly benefits from it. In reality, both motives are likely to be intertwined (Tauchen *et al.* (1991)).

[10] The increase is more acute among men than among women.

Data used in the dissertation

For **Chapter 3** of the dissertation, I had the opportunity to conduct qualitative work by doing semi-directed interviews with women in Burkina Faso. I surveyed women on marital life, child rearing, their opionion about domestic violence, and on their potential experience of domestic violence. The fieldwork has nourrished the interpretation of the results presented in the dissertation. I also took part in the collection of data used in **chapter 3**. Working on the field is a rigorous exercise, but, in my opinion, it is also one of the most enriching and rewarding aspect of the work of Development Economist.

In **Chapter 1** and **Chapter 2** of the dissertation, I use the Demographic and Health Surveys (DHS). The DHS program was established by the United States Agency for International Development (USAID) in 1984. It was designed as a follow-up to the World Fertility Survey and the Contraceptive Prevalence Survey projects. The DHS are nationally representative surveys, based on two-stage cluster sampling. Questionnaires are harmonized across countries to ease cross-country comparisons. The DHS collects data on women aged 15-49 years old. Among other topics, the DHS collect data on marriage, fertility, family planning, health including by collecting biomarkers, and on tolerance and experience of IPV. Some DHS also survey men, including some husbands of the women surveyed.

In the DHS, data are gathered on whether women believe it is justified for a husband to beat his wife trough a series of five questions: "According to you, it is justified for a husband to hit or beat his wife in the following situations:

1. if she goes out without telling him?
2. if she neglects the children?
3. if she argues with him?
4. if she refuses to have sex with him?
5. if she burns the food?"

   In addition, one eligible woman per household is randomly drawn to answer questions about domestic violence. Women are presented with situations that women can face in their every day life and are asked whether the situation apply to their relationship with their partner or husband. To elicit a controlling behavior from the partner of the woman surveyed, the following questions are asked:

(a) "Is/was your partner jealous or angry if you speak/spoke to other men?

(b) does/did he accuse you often to be unfaithful?

(c) does/did he allow you to meet with your female friends?

(d) does/did he try to limit your contacts with you family of origin?

(e) does/did he insist on knowing where you are?

(f) does/did he not trust you with money?

(g) does/did he prevent you from working?"

A short second set of questions targets emotional abuse. The measure of emotional violence boils down to two questions in the DHS:

(a) Did your husband ever do or say something to humiliate you in front of other people?

(b) Did your husband ever threatened you or someone close to you?

Based on these two questions, a dummy is defined to assess emotional violence. This variable is equal to 1 if a woman answered "yes" to at least one of the items and 0, if she answered "no" to both of them.

Then, a third part of the module, inspired from the Scale of conflict tactics developed by Straus (1990), asks whether the woman was the victim of different type of physical violence at the hand of her partner. The experience of violence is examined starting with the least severe acts of violence ("Has it ever happened that your husband slap you, shook you or threw something at you"), moving to increasingly intense forms of violence ("did he ever punch you", "did he ever dragged you on the ground or kick you?") to reach the most severe forms of physical violence ("did he ever attack you with a knife, a gun or another type of arm?"), and marital rape.

The DHS is among the most used data source on tolerance and experience of IPV, and it is undeniable that the inclusion of questions on violence in the DHS has greatly contributed to the multiplication of study on the topic, especially in the field of economics. Nevertheless, the method that laid the foundations of the design of the domestic violence module of the DHS has stirred controversy, mainly outside of the field of economics.

Measuring IPV, a critical discussion of the tools used by economists

This section aims at reflecting on the relevance of the Demographic and Health Survey (DHS) as a tool to measure domestic violence. Despite its widespread use in the economic literature, few articles provide a clear analysis of its shortcomings and of the subsequent consequences for the interpretation of the results when studying domestic violence. The DHS is the heir of the Conflict Tactics Scale (CTS) which is the leading form of quantitative survey used in the literature.

Defining violence

Defining violence against women is a central issue when studying domestic abuse as it shapes the tools used by researchers to measure it. Many of the battles that appear to be fought on the statistical field are actually deeply related to the definition of domestic violence adopted by social scientists. The debates on the relevance of the quantitative tools used to measure domestic violence are going back-and-forth between the tool itself and the definition of the phenomenon it is supposed to measure. Social scientists cannot spare the effort of expliciting their definition of domestic violence and of remaining consistent with this definition when they interpret their results.

Domestic violence against women is not uniquely defined by social scientists. As well summed up in Gordon (2000), scholars belonging to the feminist's perspective define domestic violence as a range of physical, psychological and verbal acts used to achieve domination and control over an intimate partner. To control their partner, men use a pattern of behaviors (both non-physical and physical) in the relationship (pinned down by the Wheel of Power and Control of Pence and Paymar after years of interviewing women on the field). In this approach, violence is rooted in patriarchal traditions of male dominance in heterosexual relationships. Social scientists belonging to the "conflict theory school" or family violence researchers also have studied domestic violence against women. However, both their focus and their definition of domestic violence differ strongly from that of the feminist perspective.They are interested in violence as a tactic among others (reasoning, verbal) used by family members to solve conflicts. In this approach, violence is defined as "acts carried out with the intention of causing another person physical pain or injury, regardless of whether an injury actually occurs" (Straus (1999), p. 20). As pointed out by Gordon (2000), this definition narrows down violence to an aggressive behavior that may or may not result in harm on injury. They also use the term abuse "to refer to

broad classes of physical, verbal and psychological acts intended to physically or psychologically harm or control another person in an interpersonal relationship" (Gordon (2000), p.749). The differences in the definition of violence have far-reaching implications for the understanding of the nature of violence and for our tools of measurement.

The Conflict and Tactics Scale (CTS) developed by Straus in the seventies and the controversies it triggered is a good illustration of how much definition matters. Though widely used in the literature on domestic violence, the tool raised concerns as to what it actually measures.

Quantitative measures of domestic violence against women

Quantitative measures of domestic violence are not the only tool available to study domestic violence. In-depth interviews were and still are used by social researchers. This section focuses on quantitative measures of violence because they are widespread, especially in an economic literature that has borne little interest to the strong debates that surround their use.

Victimization surveys

Victimization surveys were the first nationally representative quantitative measure of domestic abuse. They aim at measuring felonies and crimes by directly interviewing people about the violence they believe they were the victim of. The questions about violence are framed in terms of judicial category to confront it with administrative measures of crimes from police and courts (Cavalin (2010)). The US pioneered the use of large-scale victimization surveys in the seventies. The prime objective was not to assess domestic violence per se. However, a measure of violence against women, including intimate partner violence (IPV), could be inferred from them. Those surveys were developed in a context of growing concern about a potential increase in felonies and crimes in the US in the sixties (Grémy (2006)). The US launched the annual *National Crime Survey (NCS)* in 1972 and after a series of change, *the National Crime Victimization Survey (NCVS)* in 1993 (Grémy (2006)). First introduced in 1980 with face to face interviews, the survey has been completely administered by phone since 1986. Women and men are surveyed.

One of the main limitations of this type of survey was that domestic violence was subsumed within criminal categories. Additionally, framing questions in terms of crimes

may lead respondents to only declare the most violent events they went through, leading to underestimating domestic violence.

The CTS

Context and description

In parallel to the use of victimization surveys, sociologist Murray Straus developed the CTS (Straus (1979), Gelles and Straus (1979)) in the seventies. Unlike victimization surveys, Straus' CTS targets conflicts within the family. It doesn't intend to measure violence against women outside the domestic sphere. CTS was developed against the philosophical backdrop of conflict theory, which relies on the idea that conflict is an inevitable part of all human association, including family (Straus (1979)). Each member of a family has its own agenda that inevitably differs from that of other family members. What is key for the study of intrafamilial violence "is the method used to advance one's own interest; that is, the means or tactics used to resolve conflict." (Straus (1979), p.76). What will essentially differ between families is the way their members deal with those conflicts. With its check lists approach, the CTS (Straus *et al.* (1996)) gathers information on those tactics through 18 items broken down into three modes of solving conflicts: reasoning (use of rational argument and discussion), verbal agression (use of verbal and non-verbal acts that may hurt the other, threats to hurt) and violence (the use of physical violence). Though theoretically based, the distinction between the three types of tactics seem confirmed by a factor analysis (Straus (1979)). At the time, CTS was a novel instrument to measure domestic violence. The other predominant sociological perspective on the matter of domestic violence was the feminist tradition (Dobash and Dobash (1979), Stets (2012)) who studied domestic violence mainly through interviews with a sample of women in shelters or in courts, whereas the founders of the CTS aimed at developing nationally representative surveys based on it.

The CTS triggered heated debates among social scientists. In response to them, Straus *et al.* (1996) introduced a revised version of it (CTS2). Those changes include *inter alia*, more physical and psychological abuse items, additional types of sexual violence (even though it is not an obvious fit with conflict theory (Johnson (1998))) and link, to a certain extent, violent events with several types of injuries or physical outcomes it may cause. However, as noted by DeKeseredy and Schwartz (1998), introducing modifications didn't answer all the concerns voiced by CTS's critiques.

CTS's criticisms

Some of the criticisms are mainly methodological and others are more conceptual. The frontier between the two is sometimes blurry as, in the matter of domestic violence, "the discussion about the technical proprieties of the measures of violence paves the way for the debate on its sociological interpretation." (Cavalin (2013), p65).

As summed up in Dobash *et al.* (1992) and DeKeseredy and Schwartz (1998), the CTS' relevance was questioned for relying on retrospective reports of the past year's events that are unlikely to be perfectly accurate. However, this criticism could be adressed to any quantitative surveys focusing on violence against women. The CTS also came under fire for its implicit ranking of violence, from psychological to physical and sexual (a slap can make more damage than a kick, psychological violence can be more harmful than physical violence). Without systematically collecting information on the injuries or outcomes of violent acts, the distinction made in the CTS between minor and severe violence is deemed a poor operationalization of the severity of violence by Dobash *et al.* (1992). Straus (1990) already recognized those shortcomings and underlined that, as far as the categorization of violence between minor and severe violence is concerned, the acts registered as severe were the most likely to cause physical injury (Gelles and Straus (1979)).

Dobash *et al.* (1992) and DeKeseredy and Schwartz (1998) also criticized the CTS's exclusive focus on acts of violence instead of investigating as well the actor's interpretations, motivations and intentions when using the act of violence measured. According to the feminist perspective, it is problematic for two reasons. First, the CTS's ignorance of the context of violence leads to the highly controversial conclusion that there exists a symmetry in women and men's use of violence (Straus (1999)). By simply registering acts of violence without asking whether they were delivered in self-defense, the CTS appears to show that women are as violent if not more violent than men, unlike the results of victimization surveys and administrative data (police, courts, shelters).[11] Though valid, it is worth noting that the argument was anticipated by Gelles and Straus (1979): "Just because husbands and wives engage in violent and abusively violent behavior in equal numbers, this does not mean that as many husbands as wives are "abused". [...] It is important to take into account the context of marital violence. Although roughly as many wives kill husbands as husbands kill wives, studies of homicide show that wives are

[11] DeKeseredy and Kelly (1993) introduced three questions in a national study to measure the prevalence of violence in Canadian universities, to assess whether women's acts of violence occurred in self-defense and didn't find empirical support for the sexual symmetry thesis.

seven times more likely to murder in self defense (Wolfgang, 1957)." (p.27). However, even though he brings up the distinction himself, Straus doesn't fully draw the conclusions from his remark when interpreting domestic violence. For Dobash *et al.* (1992) the inability of CTS to capture the difference in nature between wife-to-husband violence and husband-to-wife violence is a testimony of the poor relevance of the CTS to measure violence against women.

Second, ignoring context leads to overestimating violence. CTS aimed at purging as much as possible the measure of violence from subjectivity. In order to do that, the tool of CTS reduces violence to a check list of named acts. However, Dobash *et al.* (1992) underlines it is illusory to believe the CTS completely erase the question of the interviewee's perception, it does not bring the rigor of observational research. Worse, they assert that the CTS inflates violence. Indeed, conflict theory conceptualizes violence as a tactic among others to settle a disagreement. This vision of violence has an impact on the framing of the survey notably through the fact that the CTS are introduced with an exculpatory statement to the respondents.[12] As a result, Dobash *et al.* underlines that researchers interpret as violent, acts that are based on respondents' admitting to acts described in such an impoverished manner that it conflates severe assaults with trivial gestures. The CTS, though framed in terms of conflict is interpreted in terms of violence (Cavalin (2013)). Dobash *et al.* (1992) gives the example of a slap. It could be a slap on the hand to chastise a dinner companion trying to reach for your dessert or a tooth-loosening assault intended to punish. "We have no more reason to suppose that people will consensually and objectively label events as instances of someone having grabbed or hit or tried to hit or used a knife (items from the CTS) than to suppose that people will consensually and objectively label events as instances of "violence" (Dobash *et al.* (1992), (p82)). This criticism can be qualified though. Indeed, the absence of context may indeed conflate a slap on the hand with the same gesture intended to hurt. But for the items "kicked", "hit or tried to hit with an object", "beat up", "choke", "threaten with a knife or gun" or "use a knife or gun" misunderstandings seem less likely. Additionally, the probability to be severely hurt is higher when "kicked", "hit or tried to hit with an object", "beat up", or "choke" than when "slapped". If the seriousness of an act of violence is defined as the relative probability that a physical act of aggression result in significant physical injury or harm,

[12] DeKeseredy and Schwartz (1998): No matter how well a couple gets along, there are times when they disagree, get annoyed with the other person, or just have spats or fights because they're in bad mood or tired or for some other reason. They also use many different ways of trying to settle their differences. I'm going to read you some things that you and your spouse/partner might do when you have an argument. I would like to tell me many times... in the past 12 months [read item] (Straus, 1990,p33)

then, even recognizing that the correlation between the two is not perfect, it can be acknowledged, with Gordon (2000), that the distinction between minor and severe acts of violence remains meaningful to operationalize the severity of the perpetrator's behavior.

In a nutshell, there is on the one hand, advocates of the CTS who see their instrument as a valid way to measure assaultive behaviors against women and whose results show a symmetry in the use of violent behavior and on the other hand, the feminist perspective who rejects the use of CTS on its own as a valid and comprehensive measure of domestic violence because it fails to capture the peculiar nature of violence against women (the pattern of control exerted by men on women and the asymmetry of this pattern).

The problematic of the discrepancies in results was tackled by Straus (1990). He relies on studies who used the CTS to survey women located in shelters and compare the frequency of their assault with women who declared they used the services of shelters in the NFVS of 1985. He finds that women registered in the NFVS who sought help from a shelter exhibit a higher frequency than women in the NFVS sample who didn't. However, the former's frequency of assaults is three times lower than the one of women in the shelter study, even using the CTS. He concludes, foreseeing its formalization by Johnson (1995) that cases of extreme violence found in shelters correspond to missing values in large scale national surveys. They are women that, regardless of the topic of the survey, are difficult to interview. He concludes that the discrepancies in results can probably be explained by a difference in the experience of violence of the populations studied rather than from the CTS's flaws. Johnson (1995) makes a similar empirical test and find similar results. He concludes: "Although I am in essentially in agreement with many of the criticisms of the CTS, [...] patterns of violence discovered in shelter samples and national samples differ dramatically even when violence is assessed with the CTS in both settings. This provides strong evidence that the differences are not due merely to the deficiencies of the CTS" (Johnson (1995), p.285). Thus the apparent inconsistencies between surveys based on CTS and those led according to the feminist perspective in shelters, comes from the fact that they survey two different and almost non-overlapping populations, who are experiencing different forms of violence.

Interpreting domestic violence with the CTS: back-and-forth between definition and measure

This last remark sends us back to the definition of violence at the core of the CTS. Indeed, whereas the feminist approach conceptualizes IPV as "a matter of control, rooted in patriarchal traditions of male dominance in heterosexual relationships, especially marriage" (Johnson and Leone (2005), p.323), supporters of the family violence theory or conflict theory (such as Murray A. Straus) see IPV as a matter of conflict that escalated to violence. Violence is defined as "acts carried out with the intention of causing another person physical pain or injury, regardless of whether an injury actually occurs" (Straus (1999), p. 20). Violence is merely a tactic used to settle a disagreement.

Johnson (1995) tries to reconcile the feminist's perspective and the results of CTS, by conceptualizing multiple forms of couple violence. Before him Steinmetz (1977) and Straus (1990) spoke about women experiencing different forms of violence. Though recognizing his distinction isn't ironclad, Johnson makes a difference between two forms of IPV: one rooted in an attempt to exert general control over the relationship (the one he calls intimate terrorism or coercive control) and the other one arising out of particular conflicts (and that he calls situational couple violence). The key difference between the forms of violence is the motivation of violence. In coercive control, a partner uses violence to control the other partner. The victims of the coercive control are overwhelmingly women whereas the common couple violence is more gender symmetric. Based on this distinction, he supports that the discrepancy in results between feminist studies and family violence advocates comes from the fact that women looking for help in shelters are very unlikely to answer to national surveys on violence against women. As a result, he expects not to be able to detect coercive control with CTS. He tests this hypothesis on a database bringing together court and shelter data and a neighborhood survey in Pennsylvania carried out in the seventies. He finds indeed a low prevalence of patriarchal terrorism (10%). Building on Johnson's distinction, Macmillan and Gartner (1999) uses the Violence Against Women Survey of Canada in 1993) and operationalize coercive control in an alternative way (using a latent class analysis instead of a principal component analysis). They find that 1.3% of women in their sample were the victim of what Johnson labels as patriarchal terrorism.

In a later article, Johnson goes further in its distinction of the types of domestic violence

in favor a categorization that transcends the conceptual frameworks used to analyze violence. According to him, common couple violence, the escalation of couple conflict to violence, is the main display of domestic violence that is recorded in national surveys. Based on samples gathered from shelters, Johnson gathered empirical evidence that supports the existence of differences in types of domestic violence.[13] Yet, this violence is less frequent and severe, unlikely to require police intervention or divorce and, more importantly, it is gender symmetric (Johnson (2006)). This last point in particular makes it conceptually problematic in a study framed in terms of patriarchy and female control by men within families. In articles published after the seminal one of 1995 though, Johnson, who first used the term patriarchal terrorism to label physical violence embedded in a pattern of control, then applied a semantic shift in favor of the term "coercive control" or "intimate terrorism". Additionally, in Johnson (2006), he comes back on its initial distinction of domestic violence and refers to it as focusing too narrowly on the behavior of one violent partner. Instead, Johnson (2006) identifies four types of domestic violence based on the behavior of both partners:

- situational couple violence: an individual can be violent and non-controlling in a relationship with a partner who is either nonviolent or violent and non-controlling;
- violent resistance: one partner is violent and non-controlling but in a relationship with a violent and controlling partner;
- intimate terrorism: a violent and controlling partner in a relationship with a partner who is either nonviolent or violent and non-controlling;
- mutual violent control: a violent and controlling partner paired up with another violent and controlling partner.

Though interesting to refine our understanding of coercive control, the classification provided cannot be implemented based on data collected in many household surveys, including the DHS.

Emancipation from the CTS

Despite its caveats, it can be agreed upon, with Gordon (2000) and Johnson (2006), that the CTS have considerably contributed to domestic violence's study by providing a standard approach to assessing violence. The debates it stirred inspired improvements in the

[13] For instance, Johnson and Leone (2005) show that intimate terrorism has a deeper and more traumatizing impact on women's well-being and health than common couple violence. Holtzworth-Munroe (2000) shows that the different types of violence refer to meaningful distinction in the profile of batterers.

design of other surveys such as the Violence Against Women Survey (VAWS) developed by Statistics Canada in the 1990's. It overcomes some short-comings of the CTS as well as some of the crime-victim surveys (Johnson (1998)). First, it is not framed in terms of conflict or dispute settlements like the CTS is. The context is clearly established for respondents in terms of experience of violence in their homes. The absence of attempt to try and normalize physical aggression should limit the risk of registering trivial gestures as violent acts as it was reproached by Dobash *et al.* (1992). A focus is put on the frequency and the consequences of the assault as well as people or organizations respondents may have tried to reach out after the attack(s) (Johnson (1998)). In the US, the National Violence Against Women Survey (NVAWS) conducted in the US between 1995 and 1996 adopts a similar definition of physical assault as the one used in the VAWS in Canada.

The VAWS and the NVAWS seem to do a better job than the CTS to measure the complexity of domestic violence. For instance, in Johnson and Leone (2005), the authors use the NVAWS to apply the distinction between common couple violence and intimate terrorism. Johnson and Leone build a control scale by using questions relative to "nonviolent control tactics" used by the partner.[14] They distinguish between low and high control of the partner, using a principal component analysis. Then, if a violent husband used fewer than 2 controlling items, the acts of violence (physical and sexual) he exerted are classified as situational violence, whereas if he used more than 3 controlling tactics, the violence exerted is labeled as intimate terrorism. Using this measure, against their expectation not to be able to capture intimate terrorism with national surveys, they find that 35% of husband's violence consists of intimate terrorism. They explain it, in agreement with Straus (1999) by the fact that the NVAW is framed as study on personal safety and has a crime overtone that may urge women who experience violence in a general context of control to report more violence than those who experience it in an isolated way. It may also be that the design of the NVAW is better suited to measure domestic violence.

Since the VAWS, surveys were designed to keep the principle of counting acts of violence but were emancipated from the framework of the CTS. It is for instance the case of the National survey on violence against women in France (ENVEFF) in 2000 or the survey Violence and Gender Relations (VIRAGE) in 2016. Both those studies focus on violence against women in general including violence in the family sphere. These two nationally

[14] These are "does your husband tries to limit your contact with your family/friends? Is he jealous or possessive? Does he insist on knowing where you are? Does he make you feel inadequate? Does he shout at you?".

representative surveys reject the conflict theory approach of violence and clearly adopt a definition close to the feminist approach in terms of domination over one's partner. They put the emphasis on the context and the severity of violence. The VIRAGE is original in that it relies on precisely described acts that are *a posteriori* qualified as judicial categories of felonies and crimes.

What type of tool is the DHS module on domestic violence?

The DHS module on domestic violence is loosely inspired by the CTS and exhibits a design rather close to the VAWS. Like the CTS and the VAWS, it relies on a check-list approach aimed at trimming as much as possible the measure of violence from the subjectivity that may come with asking women and men about the assaults or domestic violence they experienced. It uses a range of 19 questions to measure whether a woman experienced a controlling behavior from her partner or emotional, physical or sexual violence in the 12 months prior to the survey and ever. For physical violence, the questionnaire starts with questions about being slapped by the partner up to being forced to perform sexual acts they didn't want to do. The design of the questionnaire provides several opportunities to the respondent to reveal abuse. Information is also collected on the injuries women may have sustained because of the violent behavior of their partner. The frequency of violence experienced in the last 12 months is also registered but for married women only (for widows, women separated or divorced from their partners, the survey solely registers whether they were ever victim of such violence).[15] The DHS also collect information on whether a respondent was attacked by someone else than her partner. This includes people living in the same household as the respondent.

The DHS departs from the VAWS in one main aspects: its framing. Indeed, the framing of the survey is very different. If the VAWS is clearly framed as a personal safety survey, the DHS is presented as a health survey. Also, the preamble of the domestic violence module refers to a talk about relationships within a couple and to questions that can be "very personal". The respondent is also reassured about the fact that she will be the only one answering these questions and that no one will know she was asked them. So, despite the fact that it does not explicitly refers to domestic violence (contrarily to the VAWS), the framing of the domestic violence module provide sufficient hints to avoid confusing acts of violence with trivial acts. It is worth noting that questions about the tolerance of domestic violence are themselves framed (until the 2010's) in terms of conflict between partners. It might have an influence on the state of mind of the respondent when reach-

[15] sometimes=1-6 times in the year, often= 7 times or more in the year

ing questions on domestic violence. However, this hypothesis can be empirically tested taking advantage of the suppression of the preamble from 2010 on.

Main measures of domestic violence in the economic literature and ways forward

A significant share of the literature on domestic violence in developing countries, and more particularly in Sub-Saharan countries, relies on the measures of tolerance of violence and experience of violence available in the DHS.

To measure the tolerance of domestic violence in the DHS, respondents are asked whether they believe it is justified for a husband to beat his wife in five situations: whether she goes out without telling him, neglects the children, argues with him, refuses sex and whether she burns the food. Assessing the tolerance to domestic violence in daily situations is meaningful as personal exchanges, interactions and everday conflicts form "the seedbed of specific violent event [...]. [They] occur within wider cultural contexts that affect the general position of husbands and wives in a given society [...]." (Dobash and Dobash (1998), p.144). As a result, men's control and domination of women "should be discernible in the specifics of everyday life and the discourses about such relationships and their inherent conflicts." (ibid).

Studies relying on the DHS, tend to use a categorical variable for each item of the tolerance of domestic violence and/ or to gather them in a categorical variable equal to 1 if a woman answered "yes" to at least one item and 0 if she said "no" to all of them (Uthman *et al.* (2010)). Some authors also use an index scoring from 0 to 5 (Alesina *et al.* (2016), Diallo and Voia (2016), Yount and Li (2009)). Kishor and Subaiya (2008) adds to this series of measures by considering the reasons why women may refuse to have sex with their partner.[16]

Other databases than the DHS are used in the literature to assess the tolerance of domestic violence of women and men. They are sometimes used in addition to the DHS, to provide further evidence that results are consistent. Some of these measures only marginally diverge from the model set by the DHS (Yount (2005), Yount and Li (2009), Iyengar and Ferrari (2015), Koenig *et al.* (2003)). A commonly used alternative method consists in assessing how progressive women are, by asking them whether they agree or not with a series of assertions reflecting gender norms. For instance, Lowes (2016) computes an av-

[16] "A woman has the right to refuse to have sex with her husband: when she knows that her husband has a sexually transmitted disease, when she knows that her husband has sex with other women, when she has recently given birth, and when she is tired or not in the mood."

erage score based on the respondent's agreement with 18 items, on a scale of 1 (strongly agree) to 5 (strongly disagree).[17] She also uses the same 5 items of the DHS but, instead of asking it as a yes/no question, respondents are expected to say whether they agree with a statement on a scale ranging from 1 (strongly disagree) to 5 (strongly agree). Kim *et al.* (2007) also use this technique.[18] Though less common, Yount (2005) doesn't focus on precise scenarii in which wife beating is acceptable but rather asks women whether a man is never, seldom or sometimes, often justified in beating his partner.

When it comes to actual violence, the vast majority of the measures used rely on surveys whose design in close to the DHS.

Analysis conducted with DHS data either use categorical variables for each type of violence or a categorical variable equal to 1 if a woman experienced at least one of the forms of violence studied (and 0 if she experienced none of them), or build an index by counting the number of items answered "yes" by a respondent (Alesina *et al.* (2016), Pambè *et al.* (2013), Lowes (2016), Yount and Li (2009) for instance). There is little deviation from this method in the studies that build on databases other than the DHS. For instance, Kim *et al.* (2007) uses four questions to assess domestic violence, 2 on physical violence —in the last 12 months, has your partner ever pushed you or hit you with his fists or something else? —and 2 on sexual violence —in the last 12 months, has your partner physically force you to have sex when you didn't want to, has your partner have you had sex when you did not want because you were afraid of his reaction?. Then they use a categorical variable equal to 1 if any of this item was answered positively by the respondent. Bobonis *et al.* (2013) uses categorical variable to capture physical, sexual, emotional violence as well as threats of violence.

Though the CTS and NWAS-inspired surveys remain the standard tool to assess domestic violence, original measures sprung up in the literature. In Iyengar and Ferrari (2015), the authors use the HITS scale (HITS stands for Hit, Insult, Threaten, Scream) which was developed by Sherin *et al.* (1998) as a quick tool for family physician to detect abuse. This HITS scale is a 4 items questionnaire in which women are asked if their partner has hit, insulted or talked them down, threaten with harm or screamed at them. Women are given

[17] Views on gender equality average effect size (18 items): how strongly do you agree that: mother responsible for child care, man should have final say, can divorce wife if infertile, man decides when have sex, women have same right to work and study as men, women should tolerate beating, only real woman once have child, couple should decide number of children together, women can suggest condom use, men should help with household tasks, as important for girls to go to school as boys, better to have more sons, men should eat first when limited food, woman can go to health center without husband's permission, woman can use family planning without husband's permission, women should look at floor when talking to husband, wife and husband are equal partners, boys should receive education before girls if limited funds.

[18] Series of 6 statements accepting traditional gender norms, for instance, "A woman should do most of the household chores, even if the husband is not working."

multiple choices with a five point frequency format: never (1), rarely (2), sometimes (3), fairly often (4), frequently (5). The total score ranges from 4 to 20. Sherin *et al.* (1998) show a strong correlation (0.85) between the HITS score and the CTS score in the United States. Iyengar and Ferrari use it in the context of Burundi, providing a new measure of domestic violence in the African context. Eventually, in the context of the United States, Aizer (2010) had access to administrative data of women hospitalized for assault from 1990 to 2003 in the state of California. However, those data are rare and unlikely to be available in sub-Saharan countries.

Chapter 3 of the dissertation contributes to the tradition of discussion around the tools to measure IPV by assessing the relevance of an original tool: the list experiment. The innovation of the list experiment does not lie in the question asked; I still use the formulation of questions as developed in the DHS, mainly to be able to compare the results of the list experiment with results obtained with the DHS, which remains the dominant source of data on IPV in sub-Saharan Africa. But the list innovates in the protocol used to ask the question, in order to purge the measures of IPV from a social desirability bias that probably plays down the actual experience of IPV of women.

The list experiment is not the only innovative tools recently used in the literature on IPV; audio computer-assited self-interviewing (ACASI) allows the respondent to listen to pre-recorded survey questions on IPV through headphones and register their answer on their own, while the enumerator leaves the room to give her more privacy. Cullen (2020), however, finds little difference in the response rate with ACASI and with the usual questions about experience of IPV. Further research may be needed to assess the potential of ACASI. Another method whose potentiality could be explored is using vignettes.[19] Even though it was initially designed to assess the perception of a given behavior or attitude, the use of vignettes could be an interesting path to renew the way we survey women about the tolerance and the experience of violence. In particular, for experience of IPV, vignettes could be used as an entry point to discuss the respondent's own experience.

[19] Vignettes are short stories about a hypothetical person, traditionally used within quantitative or qualitative research on sensitive topics.

Outline

The dissertation is organized as follows. Each of the first two chapters explores one potential determinants of tolerance and experience of IPV: access to primary education (chapter 1), and the composition of the household (chapter 2). Chapter 3 discusses the relevance of using an alternative measure of IPV in household surveys: the list experiment. The final section concludes with directions for future research.

Chapter 1

The first chapter, co-written with Rozenn Hotte, focuses on the influence of women's own characteristics prior to marriage on their marital outcomes proxied by age at marriage, age at first child, and tolerance and experience of IPV. It zooms in on one characteristic of women: the level of education. We combine an original dataset of geolocalized primary schools and the DHS, and rely on the quasi-experimental geographical and historical variations in the number of schools built in Benin in the 1990s to analyze the impact of primary school education on women's marital outcomes.

We show that improving human capital at the primary school level has beneficial impacts on the marital outcomes of women living in rural areas and that the improvement takes place through their own human capital rather than through their husband's. We do not find evidence of a similar effect in urban areas, where women were already more educated than in rural areas before the school construction program was implemented. Using duration models to take into account the right-censored nature of the data on age at marriage and age at first birth, we show that having some additional human capital acquired in primary school postponed entry into marital life and motherhood. It is all the more interesting as the relationship is not mechanical, given that the age at the end of primary school is lower than the average age at marriage. Finally, we show that the increased human capital acquired in primary school decreases tolerance and experience of IPV.

Chapter 1 contributes to the literature on education in sub-Saharan Africa as the geolocalized dataset of schools allows us to study at a very granular level the impact of the increase in school supply on women's marital outcomes. This constitutes a substantial contribution to the education literature on Sub-Saharan Africa, that measures mainly ex-

posure to schooling at a more aggregated level. Chapter 1 also shows that, even policies that did not directly target women's marital outcomes may have beneficial impacts beyond the initial goals of the policy. Even in contexts where the quality of education tends to be low, attending primary education is nevertheless enough to shield women from experiencing IPV later in their life cycle.

Chapter 2

The second chapter explores whether having sons improve women's status in the household and shields them from experiencing domestic violence. The goal of the chapter is to explore whether a preference for sons manifests in the mother's tolerance and experience of IPV, but also in her involvment in decision-making in the household. To circumvent omitted variable bias and endogeneity issues, I use the DHS data and exploit the exogeneity of the sex of the first child and, when applicable, the fact that women are in a polygamous union to apply a husband-fixed effect strategy.

I provide suggestive evidence that the sex composition of children on the mother's tolerance and experience of IPV varies with the type of union (monogamous or polygamous), and with the rank of the spouse among women married to a polygamous partner. First wives whose eldest child is a son may benefit from more autonomy in decision-making. As for second wives, having a son is correlated with less tolerance of IPV and with an increased access to resources proxied by the probability to be anemic. I also find suggestive evidence that the longer a wife remains without a son, the more domestic violence she might experience, but that, for monogamous women, the birth of the first son is nevertheless correlated with a violent behavior from the husband. I am unable to account for the qualitative differences betwee polygamous and monogamous women. The complexity of the descriptive results also calls into question the quality of the measure of tolerance of IPV, participation in decision-making, and of experience of IPV used in large household surveys.

The paper provides an original contribution as it is at the intersection of two areas of the economic literature: the question of resource sharing within the household and more generally of relations between individuals within the household, a theme on which is superimposed the question of the determinants of the bargaining power and status of individuals, which may have an effect on domestic violence. The paper contributes to the

literature on son preference in sub-Saharan Africa by showing that having sons may have a complex impact on their mother's status and experience of IPV. Finally, the paper also contributes to the discussion on the tools used in large household surveys to measure the complex notions of status, agency, and empowerment of women in the household.

Chapter 3

The last chapter of the dissertation presents the results of a measurement exercise where I use an alternative measure of IPV, the list experiment, and compare it with the measure usually used in household surveys. The goal of the chapter is to assess the magnitude of IPV that is missed in household surveys, and to identify the characteristics of women that are missed with the usual measure of IPV.

I find that women tend to under-declare their experience of severe physical violence and marital rape by 7 to 9pp respectively when asked directly compared to what is yielded with the LE. I find no statistically significant difference when it comes to less severe physical violence. Unlike the direct measure of IPV that suggests that fewer women experience more intense forms of IPV, the LE measure yields that women seem equally exposed to the three types of IPV. The results also suggest that women's age, sex of the eldest child and exposure to a mass media campaign meant to encourage modern contraception uptake correlate differently with experience of IPV according to the method used. I notably find that mothers of first born daughter are more likely to report marital rape with the list, whereas I detect no significant effect with the direct measure.

The chapter contributes to a growing literature on measurement of sensitive behaviors by relying on data that are representative of women living in rural areas in all of Burkina Faso, rather than women living in one specific region of a country as most of the literature does. While much of the literature underlines association between women's characteristics and differences in reporting of IPV, I am able to exploit exogenous variations in the sex of the first child to investigate the potential existence of a preference for boys in rural Burkina Faso, and to study more comprehensively the impact of an edutainment program on IPV using two different measures of violence. I also provide a methodological contribution as it is the first empirical paper to include in the design of the LE safeguards suggested by econometricians that perfected the technique. Finally, the chapter may contribute to understanding the mixed results of interventions aiming at

changing gender-norms in order to decrease gender-based violence in general and IPV in particular, as they rely on an erroneous baseline account of IPV.

CHAPTER 1

EFFECTS OF PRIMARY EDUCATION ON WOMEN'S MARITAL OUTCOMES

Abstract:[1] The chapter examines the effect of human capital acquired in primary school on women's marital outcomes in Benin. We exploit a sharp increase in school constructions in the 1990s to assess the causal impact of an increase in primary school supply on primary school attendance, age at marriage and tolerance of intimate partner violence (IPV). Using the quasi-experimental geographical and historical variations in the number of schools built, we find that the school building program increased the probability of attending primary school in rural areas, increased age at marriage and decreased the probability of women being tolerant of wife beating and of experiencing IPV.

[1] The chapter is co-authored with Rozenn Hotte.

Introduction

Women's marital outcomes such as their age at marriage, their age at first child or their tolerance of intimate partner violence (IPV) have been proven to be a relevant reflection of women's well-being (Raj *et al.* (2009), Nour (2006) and Jensen and Thornton (2003)). This is especially true in developing countries, where there are few social safety net beyond the family, and where women's marital outcomes are a relevant way to gauge their status within the household.

Despite the emphasis on girls education as a tool to increase their welfare, the extent to which education can impact marital outcomes is still unclear. The literature on secondary education has found a link, but this may be somewhat mechanical notably because the time married women are expected to devote to the care of the household and of children directly competes with the time they could spend in school. In sub-Saharan Africa, being married is indeed usually mutually exclusive of going to school (Baird *et al.* (2011)). The relationship between marital outcomes and primary education however, is not automatic. Primary education could indeed postpone entry into marital and fertile life (see Osili and Long (2008) for Nigeria and Breierova and Duflo (2004) for Indonesia) but the evidence for sub-Saharan Africa remains scarce. Moreover, little has been said on the impact of primary education on another key dimension of women's marital outcomes: the tolerance of intimate partner violence (IPV).

The purpose of this paper is to look at the impact of primary school constructions on women's level of education and marital outcomes in Benin. The 1990s Beninese policy of building new schools provides a particularly relevant setting to track changes in women's outcomes, as girls' primary school's attendance had been lagging behind boys when the reform began. To shed light on this issue, we combine an original dataset of geolocalized primary schools and the Demographic and Health Surveys (DHS). We rely on quasi-experimental geographical and historical variations in the number of schools built in Benin in the 1990s. We first check whether the policy increased attendance rates, because the link between school supply and attendance is not straightforward. We then look at the extent to which women's marital outcomes were modified due to exposure to additional schools having been built and we provide extensive evidence that supports the parallel trends assumption. More precisely, we investigate the impact of the program on age at marriage, age at first child and tolerance of intimate partner violence (IPV). We interpret those changes as being due to higher primary school attendance having led

to an increase in human capital accumulation. The impact of the program is first assessed using a difference-in-difference and then applying a framework inspired by a regression kink design in order to fully exploit the variation in our data.

We show that improving human capital at the primary school level has beneficial impacts on the marital outcomes of women living in rural areas and that the improvement takes place through their own human capital rather than through their husband's. Reduced-form estimates indicate that the change in primary school supply in Benin increased primary school attendance among women living in rural areas by around 15%. We do not find evidence of a similar effect in urban areas, where women were already more educated than in rural areas before the school construction program was launched. Men who attended primary school at the same time as the women in our sample do not exhibit a significant change in primary school attendance. Using duration models to take into account the right-censored nature of the data on age at marriage and age at first birth, we show that having some additional human capital acquired in primary school postponed entry into marital life and motherhood. It is all the more interesting as the relationship is not mechanical, given that the age at the end of primary school is lower than the average age at marriage. Finally, we show that the increased human capital acquired in primary school decreases tolerance of IPV by 12 to 18% of its baseline level. Studying the pathways of the policy, we show that the results are not driven by the increase of human capital accumulation of their husbands, as they do not appear to have been impacted by the increase in school supply. Our results are consistent across methods and underline that this policy, which was not initially designed to target women's exposure to violence, had beneficial effects on females beyond the initially intended goals of policy makers.

Our results contribute to the literature in four respects. First, the geolocalized dataset of schools allows us to study at a very granular level the impact of the increase in school supply on women's marital outcomes. This gives us more statistical power to detect smaller effects compared to other papers. This constitutes a substantial contribution to the education literature on Sub-Saharan Africa, that measures mainly exposure to schooling at a more aggregated level, whether these studies focus on schooling's impact on fertility outcomes (Osili and Long (2008)) or on opinions regarding practices that are harmful to women (Cannonier and Mocan (2018)). The paper closest to ours, Cannonier and Mocan (2018), takes advantage of the variation in exposure to a free primary education program and increased funding for primary schools in Sierra Leone in 2001. They show that education reduces women's propensity to approve of wife beating. Though similar in spirit, our data allows us to deliver a more granular analysis, as we

are able to measure exposure to primary schooling as low as the DHS cluster level, whereas the above mentioned authors measure school exposure at the district level. What is more, our paper joins Cannonier and Mocan (2018) in being part of the scarce literature that addresses the causal relationship between human capital formation in primary school and tolerance of IPV in West Africa. Erten and Keskin (2018) also study study how an increase in human capital acquired during primary school impacts IPV and tolerance of it, specifically in the context of Turkey at the end of the 1990s. Like Cannonier and Mocan (2018), we find evidence that human capital acquired in primary school decreases tolerance of IPV; Erten and Keskin (2018), however, find no such effect. This difference in results is probably explained by several aspects. First, we study a policy that impacted women who were on average less educated than the women in Erten and Keskin (2018).[2] Second, the nature of the intervention we study is quite different. Erten and Keskin (2018) rely on an extension of mandatory schooling that led to an increase in the duration of mandatory primary school from 5 years to 8 years. In Benin, we rely on a policy whose main benefit was to increase primary school supply, in a context where primary education is meant to last 6 years in total and the average years of education in our sample is around 3 years. As a result, our paper documents the effects of primary education that are likely to be driven by an increase in the extensive margin of education, while the changes documented in their paper may be due to an increase in the intensive margin of primary education. Thus, both papers complement the literature as they study different contexts and interventions.

Second, we contribute to the literature by suggesting a variation in the regression kink design[3] (RKD) framework, a method rarely used so far in this context.[4] The change that we suggest in the RKD allows us to use the variation in the data more exhaustively and provide more precise estimates than with the difference-in-difference.

We also complement the literature on primary education's effects on fertility by using duration models that take into account the right-censored nature of our data. Indeed, some married women in our sample are as young as 15 years old. Girls married around this age are likely to be selected and, because their less vulnerable counterparts are not yet married, they misleadingly drive down the average age at first union in any analysis that is conditioned on being married. The use of duration models could explain why we capture an effect of primary school

[2]While 96% of women in their sample have completed their primary education, 35% of women in the sample have simply attended primary school.

[3]RKD uses a change in the slope of the probability of being treated at a kink point rather than using a discontinuity in this probability, as in a regression discontinuity design (RDD).

[4]So far, and to the best of our knowledge, the RKD has especially been used in a political economy literature on industrialized countries, in combination with administrative data (Landais (2015), Simonsen *et al.* (2010) for instance).

attendance on age at marriage and age at first child, while Cannonier and Mocan (2018) do not.

Finally, our research addresses a gap in the literature, as this study is the first to provide evidence on the causal effects of the 1990s primary school supply program in Benin. Studying this program may fuel reflections on the impacts from a series of policies based on a similar blueprint in the region, whose impact has not been systematically studied.

The remainder of the paper proceeds as follows. Section 2.2 describes the context of the increase in government spending for education in Benin in the 1990s and the data used in the analysis. Section 1.2 details the identification strategy. Section 3.6.1 presents the results. Section 1.4 discusses the potential channels that may explain the effects we find. Section 1.5 concludes the paper.

1.1 Context and Data

1.1.1 West African Education Policies in the 1990s

In 1990, 155 countries gathered at the *World Conference for Education for All* in Jomtien (Thailand), and pledged to achieve universal primary education for all children by 2015. At the end of the 1990s, this priority was reaffirmed by the international community as one of the eight Millennium Development Goals (MDGs), which world leaders designed as a framework for national policies. These two international milestones kick-started large investments in education for developing countries, including in sub-Saharan Africa, specifically in the form of school constructions and free primary schooling. At the continental level, the 1991 Conference of African Ministers of Education (MINEDAF), in Dakar, endorsed the MINEDAF VI program which launched the financial efforts needed to achieve universal education in Africa.

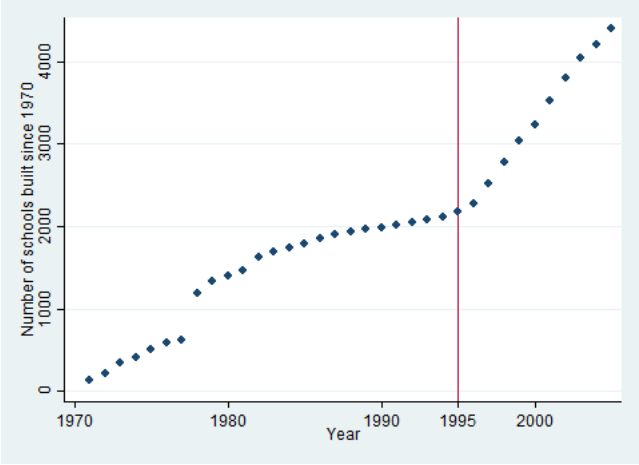
Since the 1991 Conference for Education, primary education in Benin has been promoted as a government priority.[5] The objective of one education reform launched in 1992-1993 was to improve infrastructure and increase girls' enrollment in primary school.[6] Figure 1.1 shows that the already changing trend prior to 1997 suddenly sees a sharper change in the rhythm of

[5] As presented in the National report on the Development of Education prepared for the International Bureau of Education (2001).

[6] Benin's efforts in terms of infrastructure came on top of an already existing legal framework that made primary schooling compulsory as soon as 1975. Indeed, the revolutionary regime in place at the time already considered education as a priority. Yet in 1990, at the time of a regime change and in a context where countries were pledging their commitment to a greater access to education, Benin reasserted that primary schooling was mandatory by enshrining it in the Constitution.

schools being built from that year forward.[7] Between 1997 and 2003, more than 1500 schools were built by the state or by NGOs, as can be seen in Figure 1.1. This surge reached all districts in Benin, as shown in Table A-1.1 and Figure 1.4 in the Appendix. During that period, total enrollment increased from around 722000 to 911000 pupils (26.2%).[8] The surge in the number of pupils is driven mainly by girls' enrollment and is consistent with the observed kink in the share of women who went to primary school, as shown in Figure 1.2. The figure also highlights that the policy impacted more women than men (at least for primary school attendance), since school enrollment was lower for women (1990 primary school enrollment in Benin is 27% for girls and 52% for boys[9]). The rise in school constructions starting from 1995 mirrors the steady increase in the share of women going to primary school that begins for those born after 1984. This means that even women who were 12 years old when the policy was launched are treated, which is not surprising in a context where children can enter school late and where age is not well known.

Figure 1.1: Number of schools built by year in Benin



Note: The figure presents the number of schools built by year in Benin, since 1970.

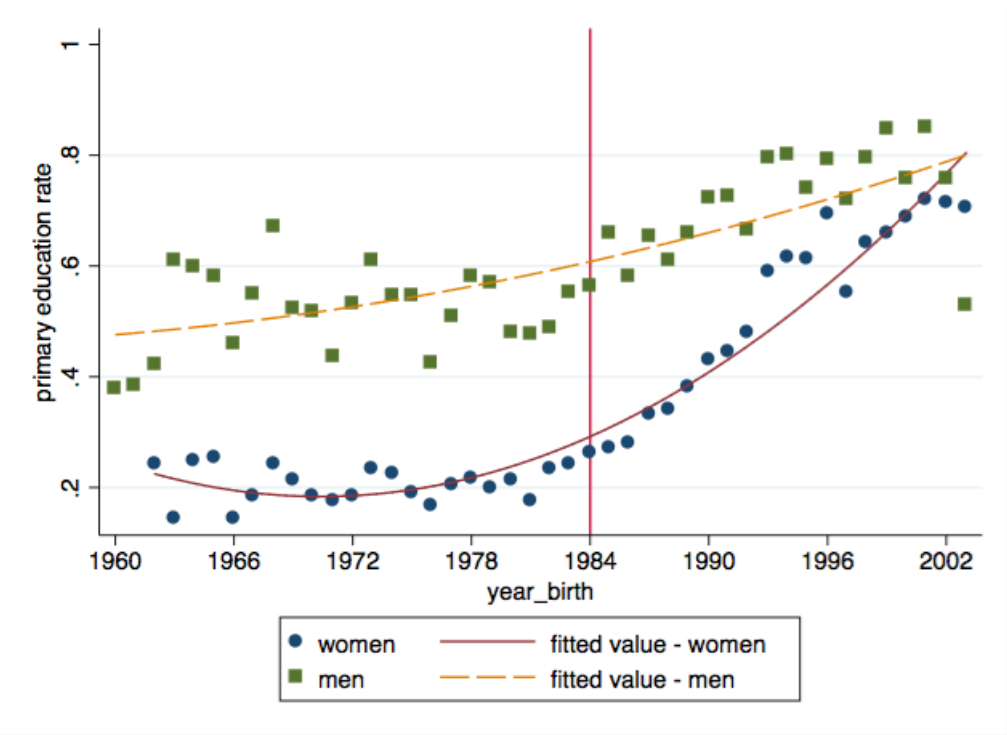
Source: PASEC data on school construction in Benin.

[7] To the best of our knowledge, there are no official documents explaining why a delay exists between announcing the policy and the change in school constructions that we see in the data. We assume that this is due, on the one hand, to potential delays in the disbursement of funds dedicated to the program and, on the other, to the time required for actually building the schools.

[8] World Bank Country Status Report: "The Beninese education system, performance and room for improvement for the education policy" (2002).

[9] These figures come from the World Bank database.

Figure 1.2: Share of women and men attending primary school by cohort in Benin



Note: The figure presents the share of women and men who attended primary school, by birth cohort in Benin.

Source: DHS Benin 2011 and 2017. Confidence intervals: 95%.

1.1.2 Data

School construction data

We use an original administrative database on school constructions in Benin.[10][11] It provides the number of schools built per year all over Benin between 1970 and 2005, and this original school construction dataset indicates the town or village where the schools were built. We geocoded the data based on the school name, for which nearly every public school is indicated with the name of the area where it is built. This allowed us to locate the schools even more precisely.

Demographic and Health Surveys (DHS)

The study we present here also relies on the DHS for Benin (2011 and 2017). The DHS collects harmonized information on women aged 15-49 years across countries and provides the geolocation of the survey clusters. Information collected in all surveys includes, *inter alia* women's marital status, age at first marriage and age at first birth, all of which we use as marital outcomes.

It also collects data on women's tolerance of IPV. The respondents are asked whether they find it justifiable for a husband to beat his wife in a series of five scenarios: if a woman goes out without telling her husband; if she neglects the children; argues with him; refuses to have sex; and burns the food. These variables are used as our main outcomes of interest and as proxies for women's empowerment and well-being. We use each binary response to the five different scenarios as outcome variables as well as an index of women's tolerance of IPV. The index is built using a principal component analysis (PCA) and retaining the first loading factor. In the 2017-18 survey wave, the survey includes a module on the experience of IPV. It is answered by a subsample of ever-partnered women. For safety reasons, one eligible women per household is randomly selected to answer the IPV module. It provides measures of the controling behavior of the partner, of emotional violence and of physical violence. Controling behavior is measured using a series of five questions: is the respondent's partner jealous if she speaks to other men, does he accuse her of being unfaithful, does he prevent her from meeting her female friends, does he try to limit her contact with her family, and does he insist on knowing where she is. We compute a score of the controling behavior of the parner with a PCA and we retain

[10] We thank Pierre André (CY Cergy Paris Université-THEMA) for having made the data available.

[11] Programme d'Analyse des systè mes éducatifs de la CONFEMEN, i.e, the CONFEMEN Programme for the Analysis of Education Systems.

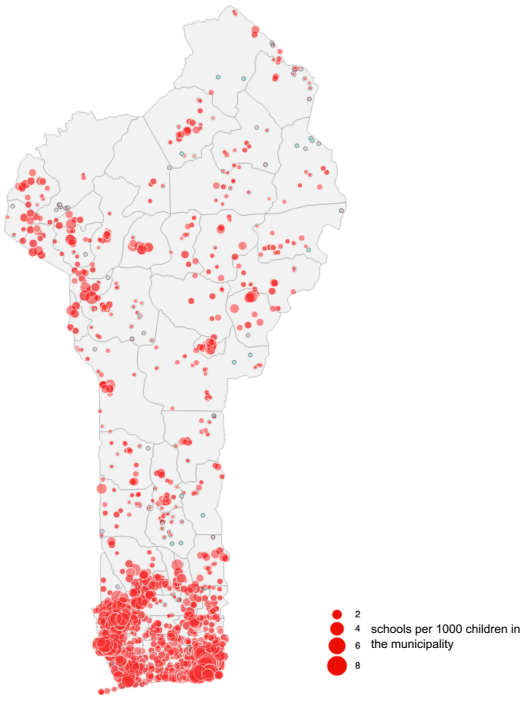
the first loading factor. Emotional violence is measured with a binary variable equal to 1 if the parner ever has humiliated the respondent, threatened the respondent with harm or insulted the respondent. Physical violence and marital rape are measured with a single binary variable equal to 1 if the respondent has ever been slapped, kicked, punched, strangled, threatened with a weapon or has ever been forced to have sex with her partner when she did not want to.

We also use women's reported involvement in the decision-making process as an outcome. In the DHS, women report who, in the household, usually decides on the respondent's healthcare, on large household purchases, on visits to family or relatives, on what to do with the money the husband earns, and on what to do with the respondent's earnings. The answers can either be the respondent alone, the husband and the respondent jointly, the husband alone, or someone else in the household. For each item, we built three binary variables to indicate whether the respondent has some say in the decision making (alone or jointly with the husband), whether the respondent was the sole decision-maker, or whether the husband is the sole decision-maker. We then built three different scores using a PCA to assess to what extent women have some say in decisions in the household, no say in decisions or to what extent they are the sole decision-maker.

Matched geolocalized datasets

We matched the school construction dataset and the 2011 and 2017 DHS datasets for Benin by projecting their respective geolocations onto a map. This allowed us to count the number of schools that had been built in a buffer around the DHS cluster of a respondent when she was of school age. In order to maintain confidentiality, the DHS Program randomly displaces the latitude and longitude of the survey clusters. They are moved by 0 to 2 kilometers in urban areas and rural clusters are displaced by 0 to 5 kilometers, with 1% of them being moved by up to 10 kilometers. Because of this random displacement rule, we build a buffer of no less than 10 kilometers in radius around the DHS clusters in order to have a measure of exposure to primary schooling that is sufficiently granular while also limiting the error in measurement induced by the displacement. It is also worth noting that the DHS Program randomly displaces clusters yet maintains them within their actual municipality. Figure 1.3 displays the number of schools built between 1997 and 2003 for 1000 children in the municipality in a buffer around each DHS clusters.

Figure 1.3: Number of schools built between 1997 and 2003 by clusters



Source: PASEC data on school construction in Benin and DHS Benin 2011 and 2017-18.

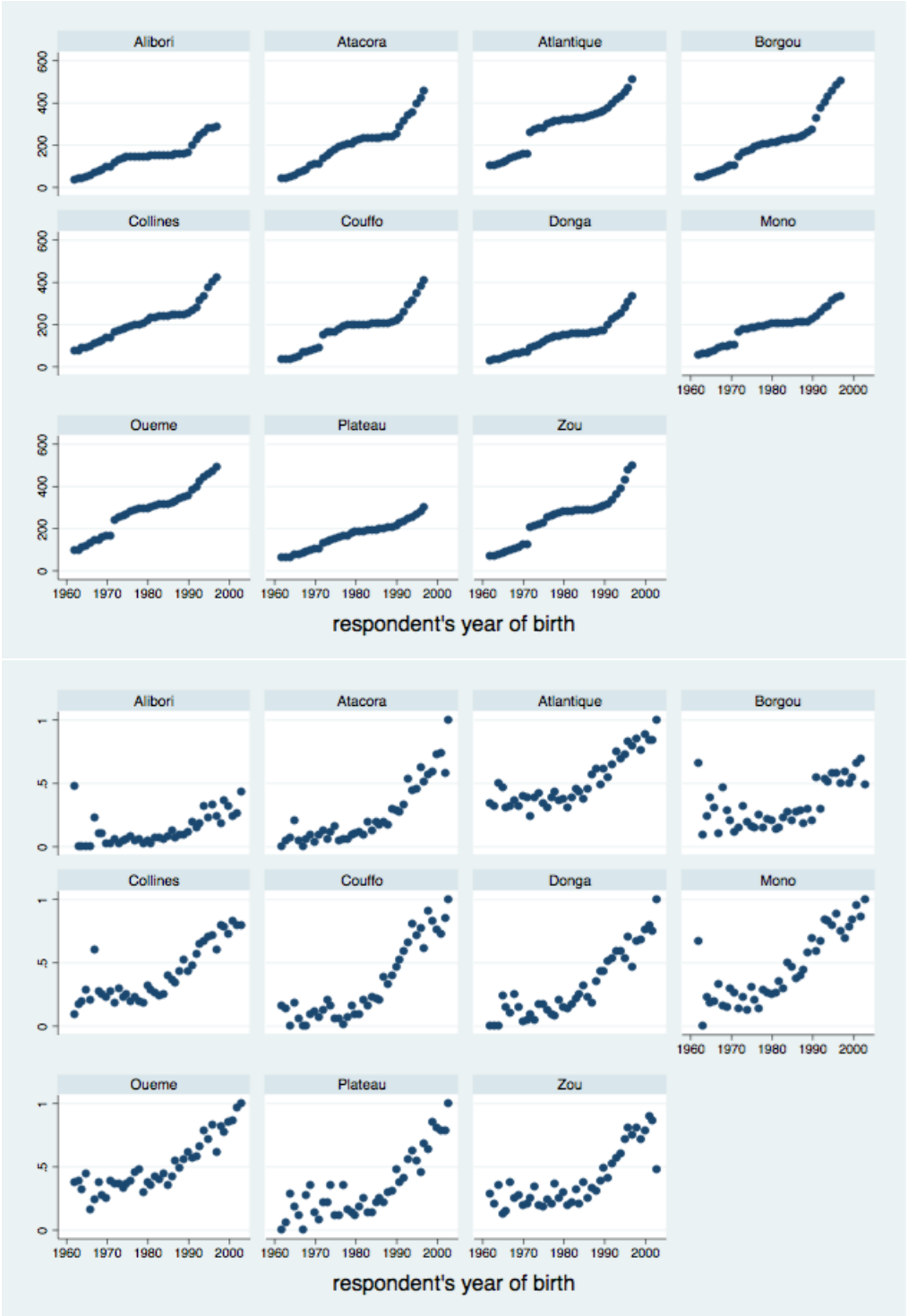
1.1.3 Geography of school constructions

We use birth cohort and geographical variation in the number of schools built as a variation of exposure to primary school. For our identification strategy to be valid, it is crucial to understand which criteria drives the construction of new schools. Our identification strategy would be threatened by omitted time-varying and region specific effects correlated with the program such as the existence of different dynamics in primary school attendance before the start of the program. To the best of our knowledge, no official documents details the drivers of the policy, so we rely on our data to identify the rationale of schools construction. Figure 1.4 describes the number of schools available in Benin when women were of school age, together with primary education by district and cohort. It shows that even though the intensity and potential response to the treatment is different across districts, the policy reached all districts in Benin.

However, as Table 1.1 and Table 1.2 show it, the program's intensity and determinants in cities appear to have been different than in the countryside, though both rural and urban areas were impacted by the school construction program.

First, the surge in school constructions relative to the initial stock of schools just before the program began, was stronger in rural areas than in urban areas. The penultimate line in Table 1.1 shows that this difference in the program's intensity is both economically and statistically sig-

Figure 1.4: Number of Primary schools and Primary school attendance by department



Source: Panel A: Data on school constructions in Benin. Panel B: DHS Benin 2011 and 2017-18.

nificant.[12] Not only was the intensity of the program different but the correlates of school constructions also differ across rural and urban areas. As shown in the last two columns of Table 1.2, the correlation between female attendance before the program and treatment intensity is six times higher in urban compared to rural areas. In rural areas, the correlation is not significant, which means that the allocation of schools is very likely to be orthogonal to our outcome of interest. [13]

It stems from these two facts that we may expect the magnitude of the effect to be stronger in rural areas, and that the validity of the identification strategy is likely to be stronger in rural areas than in urban areas. Indeed, the absence of statistically significant correlation between women's attendance to primary school before the schools were built and the locations of the schools built per 1000 children supports the hypothesis that the increase in primary school attendance in areas where more schools were built is not simply due to a catch-up effect. However, we choose to be cautious and we control for primary school attendance at the beginning of our period of analysis in our regressions. In addition, the absence of correlation between primary school attendance and the number of schools built for 1000 children makes sense as primary school attendance of girls was initially low and rather undifferentiated in rural areas.

As can be seen in Figure 1.5, before the program began, the level of education in rural areas was half as low as that of urban areas. As a result, we expect more women in rural areas to go from no schooling to some primary schooling. In other words, the policy is more likely to impact the extensive margin of primary education.[14]

[12] In this analysis, we choose to exclude Cotonou because of this agglomeration's particularity of being the economic capital.

[13] We investigate whether there are significant differences in terms of outcomes before the beginning of the program in Table A-1.2 in Appendix: there is no significant differences in rural areas. There are some significant differences for urban areas, which are not surprising since the number of schools built is much more correlated to primary schooling in this type of areas.

[14] To complete the analysis on the geographical distribution of the schools built, we investigate whether the number of schools is correlated with the share of people belonging to a specific ethnic group, or religion. The number of schools built by cluster is not correlated with the share of Christians or Muslims in the cluster. It is positively correlated with the share of Fon and Adja, which are the major ethnic groups in Benin (the Fon represent 37% of the population and the Adja 15%). It is not surprising since people from these ethnic groups live more in the south where more schools were built. We will control for the ethnic group in all our regressions.

Table 1.1: Number of schools built

Variables	Urban	Rural	Diff.
Number of schools in the cluster in 1996	2.52	1.96	0.56***
			(0.00)
Number of schools built in the cluster between 1997 and 2003	18.89	12.55	6.33***
			(0.00)
Number of schools built in the cluster between 1997 and 2003 for 1000 children	1.08	0.91	0.16***
			(0.01)
Number of schools built between 1997 and 2003/Stock in 1996	0.55	0.67	-0.13***
			(0.00)
Number of clusters	413	719	1132

Note: The table presents the differences in mean of school allocations between 1997 and 2003, according to the status of the cluster (in a rural or urban area). We look at the stock of schools available in 1996 in the cluster, before the policy, and to the number of schools built between 1997 and 2003, in absolute terms and in relation to the already available stock. Significance levels are denoted as follows: * p<0.10, ** p<0.05, *** p<0.01. Sample: DHS clusters. Cotonou is excluded.

Source: DHS Benin 2011 and 2017.

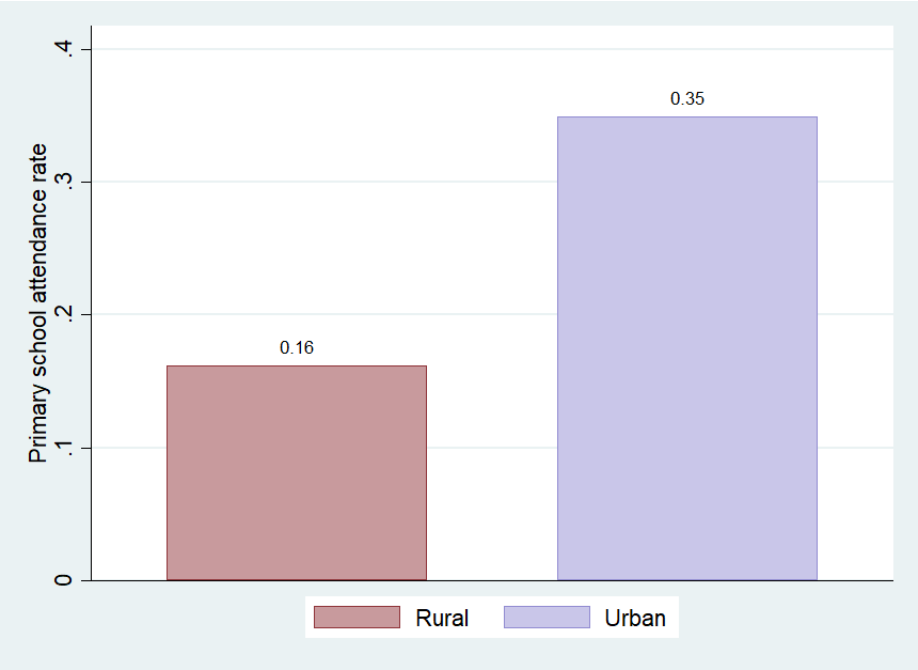
Table 1.2: Allocation of schools between 1997 and 2003

	Number of schools		Number of schools for 1000 children	
	Urban	Rural	Urban	Rural
Number of children in the municipality (thousand)	0.858***	0.256***	-0.017***	-0.034***
	(0.08)	(0.07)	(0.01)	(0.01)
Female primary attendance average	9.576***	1.240	0.598***	0.142
	(2.25)	(1.82)	(0.16)	(0.13)
Number of clusters	407.00	713.00	407.00	713.00
r2	0.32	0.02	0.04	0.05
F	97.07	6.04	8.67	18.89

Note: In the first and the 4rth columns, the average of each outcomes at the level of the cluster is displayed, for areas where the number of schools built for 1000 children is higher than the median (respectively for urban and rural clusters). In the second and the fifth columns, the average of each outcomes at the level of the cluster is displayed, for areas where the number of schools built for 1000 children is lower than the median (respectively for urban and rural clusters). The third and the 6th columns display the results of the ttests. Significance levels are denoted as follows: * p<0.10, ** p<0.05, *** p<0.01. Sample: DHS clusters. Cotonou is excluded.

Source: DHS Benin 2011 and 2017.

Figure 1.5: Primary School Attendance rate among women aged 12 and more in 1997



Source: DHS Benin 2011 and 2017. Women born between 1980 and 1985. Cotonou is excluded.

1.1.4 Raw difference-in-difference

Before moving to the empirical approach, we first provide descriptive evidence of the effects of the school constructions on primary school attendance in rural areas. We split clusters according to the intensity of school constructions: "high-intensity" zones are defined as the clusters that received more than the median number of schools built in rural areas, and "low-intensity" zones are the other ones. We split the individuals according to their age at the time of the boost in school constructions, with people aged 4-8 the ones who could have benefited from primary school constructions. The first column of Table 1.3 presents the average primary school attendance rate for women in the high-treated areas. The second column presents this average for individuals in the low-treated areas. The third column presents the difference between high intensity areas and low intensity areas for a given age group, and the last column shows the difference of the differences between the two age groups. Table 1.3 arguably suggests that people of schooling age at the time of the sharp increase in school constructions and residing in high-intensity areas are 13 pp more likely to have attended primary school than people of the same age in low-treated areas. In contrast, this gap remained steady around 3 pp for older generations. Even though there are some initial differences in the levels of primary education between high-intensity and low-intensity areas in the older placebo age groups, the age-specific trend in primary education in the placebo groups is constant. It provides descriptive evidence that we are unlikely to confound the impact of the treatment with pre-existing drivers of change. Fig-

Table 1.3: Primary education

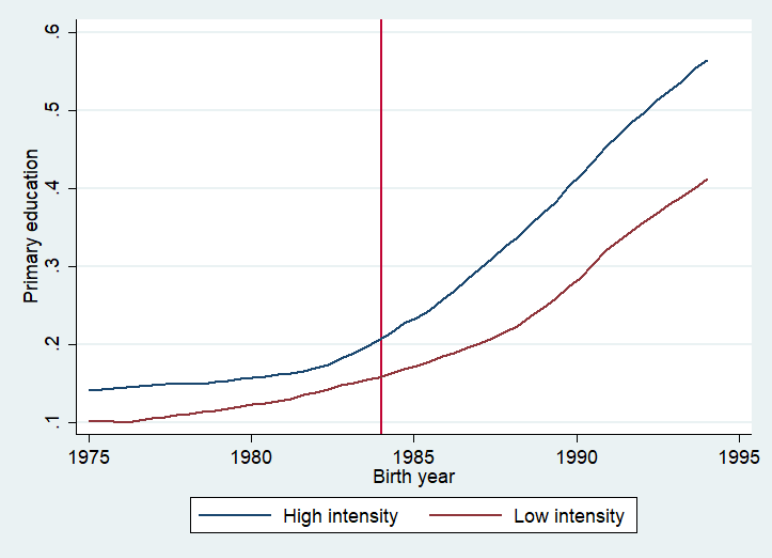
	High intensity	Low intensity	Diff	
Ind. aged 4-8	0.45	0.32	-0.13***	
Ind. aged 12-17	0.18	0.15	-0.03**	
Double difference				0.10***
Ind. aged 12-17	0.18	0.15	-0.03**	
Ind. aged 21-26	0.14	0.10	-0.04***	
Double difference				-0.01

Note: The first column presents the average primary school attendance rate for women in the high-treated areas. The second column presents this average for individuals in the low-treated areas. The third column presents the difference in average, and the last one the double difference. Significance levels are denoted as follows: * p<0.10, ** p<0.05, *** p<0.01.

Source: DHS Benin 2011 and 2017. Women residing in rural areas.

ure 1.6 further supports the claim that there was no differential pre-trends in primary school attendance between high and low intensity areas and Figure 1.7 suggests it is also the case when looking at tolerance of domestic violence.[15] Now, we move to establish a causal impact of the school constructions.

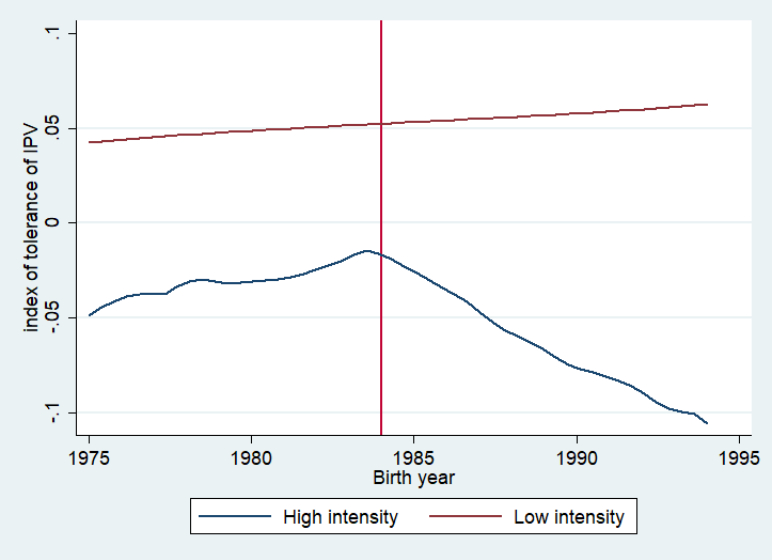
Figure 1.6: Primary School Attendance by high/low intensity regions and birth cohorts in rural areas



Note: the figure represents the level of primary school attendance by high/low intensity areas and birth cohorts smoothed with a Kernel-weighted local polynomial. **Source**: DHS Benin 2011 and 2017. Women born between 1974 and 1995 residing in rural areas.

[15] We do not present here the same figure using age at marriage, due to the right-censorship of the variable.

Figure 1.7: Index of tolerance to domestic violence by high/low intensity regions and birth cohorts in rural areas



Note: the figure represents the level of the index of tolerance to domestic violence by high/low intensity areas and birth cohorts smoothed with a Kernel-weighted local polynomial. The Index is built using a PCA analysis. **Source**: DHS Benin 2011 and 2017. Women born between 1974 and 1995 residing in rural areas.

1.2 Methodology

1.2.1 Difference in Difference

Specification

Following in the footsteps of Duflo (2001), we first use a difference in difference to identify the causal impact of the rise in school constructions on our outcomes of interest. We exploit the fact that women's exposure to the policy varies according to their birth cohort and municipality of residence. We are more precise as we take advantage of the DHS clusters having been geolocalized in the 2011 and 2017 DHS surveys. We measure exposure to the program as the number of schools built between 1997 and 2003, within a 10 km radius around each DHS cluster of residence of women. As a result, exposure to the program varies according to a woman's place of residence and age at the time the program was implemented. We proxy women's place of birth with their place of residence. The implications of this approximation are discussed in Subsection 1.3.2. Age at entry into primary school is set to 6 years old in Benin but we cannot completely exclude late entry or cases of remaining in school beyond the official age. With that in mind, our first specification defines the exposed cohort as women aged 4 to 8 years in 1997[16] and the untreated cohort as women aged 12 to 17 years when the program began. This choice of cohort may lead to an attenuation bias as some women in the control cohort may have been exposed to the education program due to late entry into school. We estimate the following model:

$$y_{imcg} = a_0 + \beta_c + \theta * N_g + \delta * N_g * TREAT_i + \alpha_m + \eta X_i + \gamma Z_{mc} + \varepsilon_{imcg} \tag{1.1}$$

where y_{imcg} is the outcome of interest for the woman i residing in municipality m and born in year c; a_0 is a constant; α_m is a municipality of residence fixed effect;[17] β_c is a birth cohort fixed effect and N_g is the number of schools built between 1997 and 2003 in a 10 km radius around a woman's DHS cluster of residence. This can be read more broadly as the intensity of the program in cluster g. $TREAT_i$ is a dummy variable equal to 1 if the woman was born between 1989 and 1993 (meaning whether the woman was between 4 and 8 in 1997). It is equal to 0 if she was born between 1980 and 1985 (meaning whether the woman was between 12 and

[16] A UNESCO report published in 2014 (Equipe nationale du Bénin and de Dakar (2014), p.81) establishes that among children attending the first year of primary school in 2009-10, 92% of them were between 5 and 8 years old.

[17] We do not apply a DHS cluster fixed effect because the number of observations in each cluster ranges from 7 to 42 at an average of 23, which we deem to be too few.

17 in 1997).[18] We also add a set of individual controls X_i, which include religion and ethnicity. Finally, Z_{mc} is a municipality-specific year effect for the density of school age children before the program began. This particular control is added because, should the density of school-age children play a role in Beninese school construction, we believe that the impact of the initial density may vary over time according to municipality. As we are using several waves of DHS, we are able to control for age effect by introducing the age and the age squared in the specification. Including these controls is a way to limit the noise brought by age heaping, which is very common in this context. Standard errors are clustered at the DHS cluster level, which is the level at which we measure our treatment.

To identify precisely which cohorts is impacted by the program, we also present in Section 3.6.1 the effect of the policy on the outcomes of interest per age at the time the policy was implemented which is tantamount to studying the policy's effect on a given cohort. The results shown are yielded by the following specification:

$$y_{imcg} = a_0 + \beta_c + \theta * N_g + \sum_{a=2}^{21} \delta_a * (N_g * v_{ia}) + \alpha_m + \eta X_i + \gamma Z_{mc} + \varepsilon_{imcg} \tag{1.2}$$

where y_{imcg} is the outcome of interest for individual i, residing in municipality m and born in year c; a_0 is a constant; α_m is the municipality of residence fixed effect; β_c is a birth cohort fixed effect; N_g is the number of schools built between 1997 and 2003 in a 10 km radius around a woman's DHS cluster of residence g; v_{ia} is a dummy indicating whether individual i was age a in 1997; X_i includes the individual's religion and ethnicity, and Z_{mc} is a municipality-specific year effect for the density of school-age children before the program began. Standard errors are clustered at the DHS cluster level.

Identifying assumptions

The key identifying assumption to interpret a causal effect is that, in the absence of the boost in school constructions, the increase in primary school attendance would not have systematically differed between areas with different exposure to the treatment. We already showed in section 1.1.1 that the localization of schools was not correlated with the attendance rate before the program began, and more importantly, was not correlated with the pre-trend in primary school attendance in rural areas. However, since the intensity of the treatment correlates with

[18] Because of late entry into primary school, some of the women born in 1985 could have been exposed to the treatment. If this were the case, it would bias our estimates downwards.

the initial attendance rate in urban areas, we also control for the interaction between cohort of birth dummies and the initial attendance rate at the municipality level.[19] When presenting our results, we also provide placebo tests using Model 3.3. In those placebo tests, $TREAT_i$ is a dummy variable equal to 1 if the individual was born between 1980 and 1984, and it is equal to 0 if she was born between 1971 and 1976. In other words, our placebo compares women aged 13 to 17 years in 1997 to women aged 21 to 26 years in 1997. Those women are not expected to have benefited from the education program. If their education had started to increase in treated regions before the surge in school construction, the coefficient δ would be positive and significant.

The identification assumption would also be violated if they were omitted time-varying and regional specific effects correlating with school constructions such as other governmental programs initiated at the same time.[20] To our knowledge, no other contemporaneous program was implemented at the same time. This is not surprising, since education had been identified as a priority and funding was limited. Furthermore, since we are exploiting the geographical variation in schools construction at a very granular level, it seems highly unlikely that we are capturing the impact of any other campaigns.

Migration is another concern for the identifying strategy; as the DHS data does not provide the place of birth for the two waves, we assess the impact of the program using the number of schools built in the respondent's place of residence rather than the number of schools built in her place of birth when she was of schooling age. This approximation can induce measurement errors or endogeneity problems, if schools constructions impact migration for instance. Even though migration is very limited in this context, as only 18% of women born between 1989 and 1993 have migrated between municipalities, we check whether the stability of our results to accounting for migration in section 3.6.1 of the paper. We notably use data on migration collected in the 2017-18 DHS survey wave, indicating whether the respondent has always been living in her current residence and rerun the analysis on women who have never migrated. We find consistent results on this subsample. We also test whether several variables including school expansion predicts migration.

[19] As suggested by Duflo (2001), it also allows us to ensure that our estimates do not capture a simple reversal to the mean for the primary attendance rate (and therefore a difference in pre-trends).

[20] In order to incentivize parents to enroll girls to primary school, fee waivers were decided for them in 1993 but the measure was made effective in 2006 Gastineau *et al.* (2015).

Rationale for focusing on reduced-form evidence

In Section 3.6.1, we focus mainly on the interpretation of reduced-form estimates. As we are interested in knowing the impact of primary education on women's marital outcomes, we could be tempted to instrument primary education by school constructions. Though our interpretation is that exposure to treatment leads to an increase in human capital through higher attendance, we are cautious when interpreting the two-stage least square estimates for the impact of primary education. Our reasoning for this is as follows.

First, there exists a potential simultaneity bias. Indeed, nearly 15% of the women in our sample marry before 15 years old. Because of late entry into primary school, it is reasonable to expect that some parents make a simultaneous rather than sequential decision on whether to marry their daughter or keep her in school. Rosenzweig and Wolpin (2000) have shown that even when using a natural experiment, this simultaneity bias prevents researchers from pinning down causal estimates. When translated to our context of studying the impact of the education policy on marriage, instrumenting education with exposure to school constructions could potentially violate the exclusion restriction.

Second, even though we expect women to marry after they leave primary school, an increase in school constructions could impact marital outcomes because the education policy may also spur changes in the norms of age at marriage or tolerance of domestic violence without these changes having been caused by a girl's own education. In this case too, the exclusion restriction would be violated. We are not agnostic about the channel through which the policy impacts marital outcomes, as we believe it is driven by women's accumulation of human capital in primary school, whether it be a woman's own capital accumulation or that of her peers. Finally, we also err on the side of caution by focusing on reduced forms, as the educational program could have simultaneously impacted the quality of education.[21]

1.2.2 Regression Kink Design

Specification

Emulating Duflo (2001) by using a difference in difference has become a classical method in the development literature for assessing the impacts of a shock whose effect varies across time and space. Yet, given that the increase in school constructions in Benin in the 1990s follows a linear trend and that our current specification absorbs a good share of this variation with the

[21] The negative impact of school expansion programs on the quality of education is well documented in many contexts (Duraisamy *et al.* (1998), Deininger (2003)).

birth cohort fixed effects, we choose to apply another strategy, inspired by a regression kink design (RKD). We call this strategy "kink in difference", and this design, allows us to more fully exploit the geographical and the time variation in the data.

Originally, the RKD exploits a change in the slope of the likelihood of being treated at a kink point. If the outcome also exhibits a kink at the same point, then the causal impact is found by dividing the change in the outcome slope by the change in the treatment slope. This method has often been used in public economics (Simonsen *et al.* (2010), Landais (2015), Card *et al.* (2012) and Card *et al.* (2015)). It allows us to use the information included in the slope of the treatment for every individual born around the kink. Since the program triggered a change in the trend of the number of schools being built continuously over time, the approach seems to be also suitable for this setting.

Here, we draw inspiration from the RKD and exploit the change in trend of exposure to schooling, which we define as the number of schools built between 1997 and 2003 in the 10-kilometer radius around a given cluster and according to birth cohort. 1984 is chosen as the kink point, based on Figure 1.2.

Since we use both historical and geographical variations, we will look at the following reduced form:

$$\begin{aligned} SchoolAttendance_{img} = a_0 + \alpha_m + \beta * (BirthCohort_i - 1984) + \gamma * (BirthCohort_i - 1984) * Post + \\ \lambda * (BirthCohort_i - 1984) * Post * N_g + \theta * N_g + \\ \mu * X_i + \varepsilon_{img} \end{aligned} \tag{1.3}$$

where N_g is the number of schools built between 1997 and 2003 in a 10 km radius buffer around individual i's DHS cluster of residence g, and α_m is a municipality of residence fixed effect. The coefficient of interest is λ, which measures the change in the slope of school attendance, by municipality, once the policy has been implemented. We also add municipality fixed effects and individual controls (religion and ethnicity), as well as differentiated trends according to the initial enrollment in the municipality and the initial number of school-age children. Instead of a birth cohort fixed effect, we include a time-trend control. Since the increase in exposure to schooling is linear, this strategy is more flexible than using a difference in difference specification with birth cohort fixed effects. It also provides more statistical power. Standard errors

are clustered at the DHS cluster level. We show in Section 3.6.1 that the two strategies yield consistent estimates.

RKD's identifying assumptions

There are several identifying assumptions that need checking in order to ensure the validity of the method. First, a kink must exist in the probability of being treated (i.e., the exposure to the number of schools built) according to the running variable, which is the year of birth. This is highlighted in Figure 1.1. Second, the treatment assignment must be "as good as random" at the threshold, i.e., the kink point. Following the literature, we thus check whether covariates are smooth at the kink. As highlighted in the Table A-1.3 in the Appendix, covariates exhibit no kink at the threshold. In our opinion, it is unlikely that the women in this specific case intentionally manipulate the running variable that is to say their age. In addition to that, women at the time of survey had neither interest in nor any reason for declaring that they were born before or after 1984. Despite this, we use a density test to check that there is no manipulation of the running variable at the kink: we perform a McCrary test and find no change in the density at the kink, as highlighted in Table A-1.4 of the Appendix.

1.2.3 Duration Model of Entry into Marriage or Motherhood

We identify the effect of primary education on women's well-being in a sample of women aged 18 to 32 years at the time of the survey. Yet, the median age at marriage in Benin is nearly 18 years old. As a result, there is a non negligible share of right-censored observations when we study marital and motherhood outcomes (age at first marriage and age at first child). As a consequence, the difference-in-difference strategy (or the RKD) for these outcomes yields estimates that are biased by women who entered their marital or fertile life earlier than the average Beninese woman. To circumvent this selection issue, we use a duration model of entry into marriage and into motherhood. The duration models are able to deal with right-censored observations in ways the usual regression models cannot. Such models have been used in the literature to pin down socio-economic correlations with birth spacing in Sub-Saharan Africa (Ghilagaber and Elisa (2014)) and to study preference for sons through birth spacing (Lambert and Rossi (2016), Rossi and Rouanet (2015)).

We use a discrete-time duration model to test whether being exposed to more primary schooling is related to a delay in marital and fertile life. Though it is rather common in the literature to use either a proportional hazards (PH) model or a Cox model, we choose the discrete time

duration model for two main reasons. First, in models such as the Cox model, time is strictly continuous and there can be no simultaneous events. A duration, t_i, that leads to the studied event should be associated with one observation, i, if the clock for measuring the duration is sufficiently precise. Yet the DHS (as well as many other household surveys), collects time-discretized data. As a result many simultaneous events occur (women born the same year entering their first unions at the same age, for instance) which violates a necessary condition of the Cox model.[22] Second, the Cox model relies on the assumption of PH, which in our case translates into the ratio of the risk of experiencing the event being constant between treated and untreated women at every moment within the studied duration.[23] Yet, we could imagine that, among the women from older cohorts, the risk of experiencing the event of interest (first marriage or first birth) intensifies at an earlier moment in the studied duration than it will for treated women.[24] The discrete-time model allows us to circumvent this potential issue since time is introduced as a covariate. The risk is modeled as a conditional probability and the estimation relies on the maximization of a binomial-type likelihood. The most commonly used function is the logistic regression:

$$\begin{aligned} log(\frac{p_{timcg}}{1-p_{timcg}}) = a_0 + a_k * t + a_p k * t^2 + \alpha_m + \beta_c \\ + \theta * N_g + \delta * N_g * TREAT_i + \gamma X_{mc} \end{aligned} \tag{1.4}$$

where p_{timcg} is the probability of experimenting the event and t is the number of years passed since the respondent's birth. [25] In the double-difference approach, X_{mc} includes N_g as the number of schools built between 1997 and 2003 in a 10 km radius buffer around an individual DHS cluster; $TREAT$ is the binary exposure to treatment according to the birth cohort; and α_m is the municipality of residence fixed effect. Individual controls (religion and ethnicity) are also included, as well as birth cohort fixed effects and their interaction with the initial attendance rate and the density of school-age children at municipality level.

For the RKD-inspired specification, X_mc includes N_g as the number of schools built between 1997 and 2003 in a 10 km radius buffer around an individual DHS cluster of residence, and the

[22] This constraint can be alleviated by correcting the partial likelihood function with simultaneity using the "Breslow" method, which is used by most statistical software and programming languages, such as STATA and R, respectively.

[23] It is important to remember that the duration studied corresponds to the years between the respondent's birth and her first union/child.

[24] The log-log plot test seems to suggest otherwise, however, which points to the respect of the PH assumption.

[25] For this analysis, the data is reshaped: there is one observation per year and per woman until she get married or, if she does not, until she is surveyed.

birth cohort is centered at the kink $BirthCohort_i - 1984$, which interacts with $Post$, $(BirthCohort_i - 1984) * Post * N_g$. X_{mc} also includes a municipality fixed effects and individual controls (religion and ethnicity).

1.3 Results

1.3.1 School Attendance

Double Difference

All the tables in this section are divided into *Panel A*, which shows the results of the regressions of interest, and *Panel B*, which displays the results of placebo regressions. The placebo difference-in-difference regressions rely on the comparison of two cohorts that are both presumably unaffected by the policy intervention. In *Panel A* of Table 1.4, we provide evidence that the education policy increased primary school attendance among school-age girls in 1997. On average, one school built per 1000 children in a 10 km radius around a cluster in a municipality increases the probability of primary school enrollment by 4.1 percentage points in rural areas.[26] The effect is robust to the inclusion of controls for school-age children's enrollment and density in the municipality before the program began. The placebo test in *Panel B* shows that earlier cohorts were, as expected, unaffected by the policy. The test also hints at the fact that there was no pre-existing change in the primary education trend that may be confounded with the school program's effect.

Figure 1.8 singles out the cohorts that were most affected by the program and shows the coefficients identified by equation 3.4. The figure provides visual confirmation that girls in rural areas aged 4 to 11 in 1997 benefited from the program, and it further suggests that the younger they were in 1997, the more intense was the effect of school constructions on primary school attendance. The sharp drop in primary school attendance for people aged 17 and 12 in 1997 are reflecting sharp patterns of age heaping, which is a well-known phenomenon in many sub-Saharan African countries. These errors can notably manifest as a result of a preference for round numbers in a context where ages are precisely reported.[27] While the age heaping pattern manifests every 5 years, the underlying trend moves up overtime. This age heaping adds noise to our estimation but does not prevent us from identifying a clear change in the trend for

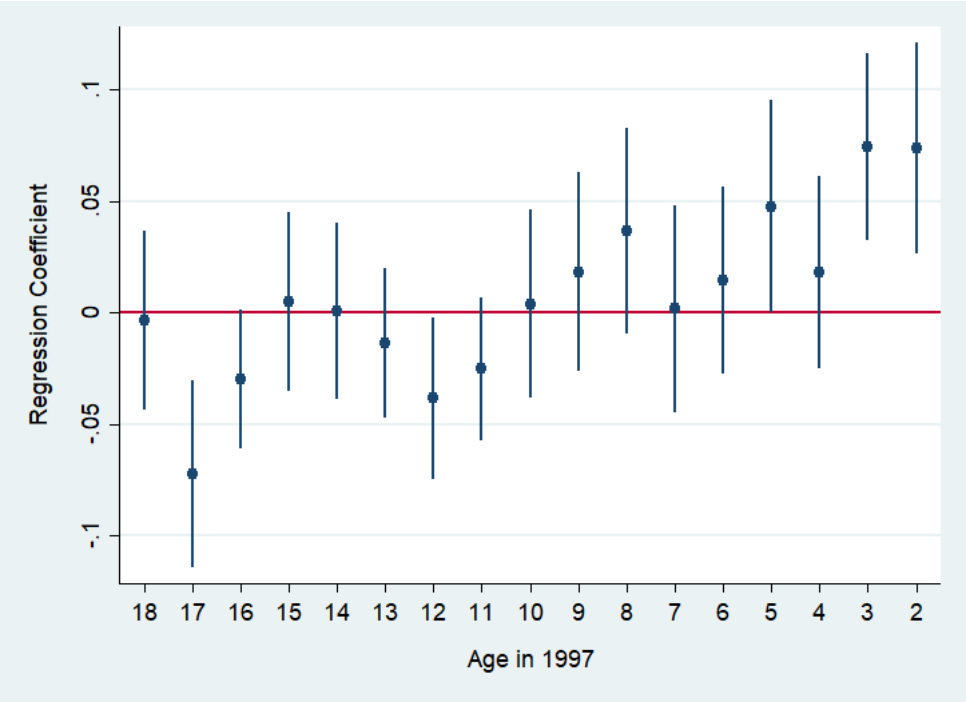
[26] When a child born in 1982 was ten years old, there were on average 1.2 schools per thousand children. For a ten-year-old child born in 1992, they were on average 1.9 schools per thousand children

[27] This is not an issue that arises in similar studies on other continents like for instance in Indonesia (Duflo (2001)) or in Turkey (Erten and Keskin (2018)).

beneficiaries of the school construction program. Table 1.5 provides evidence that the school construction program increased years of schooling among rural girls by 0.2 years. The modest change in years of schooling suggests that the program mainly impacted the extensive margin of primary education.

Since we are interested in education-induced improvements in women's well-being, our analysis from this point forward will focus on women currently living in rural areas.[28] Table A-1.5 in the Appendix shows that our results on education are robust to the inclusion of younger cohorts.[29] Bearing in mind that we capture effects only in rural areas and no effects in urban areas (where growth may be higher), this suggests that it is unlikely that a change in growth trend drives our results.

Figure 1.8: Effect of the treatment on school attendance by birth cohort in a context of strong age heaping



Note: The figure presents the coefficients of the interaction of respondent's age in 1997 and the number of schools built between 1997 and 2003 in the region of residence in equation (2). The dependant variable is having attended primary school. Sample: Rural women aged 15-49 years old.
Source: DHS Benin 2011 and 2017.

Table A-1.6 in the Appendix indicates that the intervention's benefits for primary school at-

[28] As shown in Section 2.2, schools allocation is orthogonal to prior primary school attendance of women, whereas it is not the case in urban areas.

[29] It is also possible that the effect is not significant and has a negative sign in urban areas due to the way we approximate the number of schools built in the current place of residence. Many more migrant women live in urban areas (among the treated women 30% versus 16% in rural areas), and this high level of migration could introduce complex bias in the results. This provides an additional reason for being more confident about the causal impact of the treatment for rural women.

Table 1.4: Probability of primary school attendance

	All	Urban	Rural
Panel A: Interest Experiment: Individuals aged 4 to 8 or 12 to 17 in 1997			
Number of school built between 1997 and 2003 * Treat	0.018	-0.004	0.041***
	(0.01)	(0.02)	(0.01)
Controls Individual	Yes	Yes	Yes
Mean Dep. Var.	0.35	0.47	0.27
N	10026	3785	6241
r2	0.25	0.27	0.21
F	24.20	13.27	13.93
Panel B: Placebo Experiment: Individuals aged 13 to 17 or 21 to 26 in 1997			
Number of school built between 1997 and 2003 * Placebo	-0.003	0.003	-0.006
	(0.01)	(0.02)	(0.01)
Controls Individual	Yes	Yes	Yes
Mean Dep. Var.	0.21	0.33	0.14
N	7947	2830	5117
r2	0.23	0.29	0.11
F	7.93	5.38	2.53
Regression Kink Design			
Normalized birth year*			
post kink*	0.003**	0.000	0.006***
Number of schools built			
	(0.00)	(0.00)	(0.00)
Controls Individual	Yes	Yes	Yes
Mean Dep. Var.	0.33	0.46	0.25
N	17680	6502	11178
r2	0.25	0.25	0.20
F	73.55	26.94	47.76

Note: The dependent variable is having attended primary school. The panel A and panel B present the results of the double difference model. We control by the ethnicity and the religion of the woman. Significance levels are denoted as follows: * p<0.10, ** p<0.05, *** p<0.01. Sample: Rural women aged 15-49 years old.

Source: DHS Benin 2011 and 2017.

Table 1.5: Number of years of education

	All	Urban	Rural
Panel A: Interest Experiment: Individuals aged 4 to 8 or 12 to 17 in 1997			
Number of school built between 1997 and 2003 * Treat	0.111	0.006	0.202*
	(0.11)	(0.17)	(0.12)
Controls Individual	Yes	Yes	Yes
Mean Dep. Var.	2.44	3.65	1.71
N	10026	3785	6241
r2	0.27	0.28	0.22
F	24.61	13.14	13.32
Panel B: Placebo Experiment: Individuals aged 13 to 17 or 21 to 26 in 1997			
Number of school built between 1997 and 2003 * Placebo	0.019	-0.021	0.025
	(0.09)	(0.17)	(0.09)
Controls Individual	Yes	Yes	Yes
Mean Dep. Var.	1.22	2.14	0.72
N	7947	2830	5117
r2	0.23	0.26	0.11
F	7.45	5.39	2.41
Regression Kink Design			
Normalized birth year*			
post kink*	0.023*	0.005	0.040***
Number of schools built			
	(0.01)	(0.02)	(0.01)
Controls Individual	Yes	Yes	Yes
Mean Dep. Var.	2.23	3.40	1.54
N	17680	6502	11178
r2	0.28	0.27	0.22
F	71.11	32.74	38.38

Note: The dependent variable is the number of years of education. The panel A and panel B present the results of the double difference model. We control by the ethnicity and the religion of the woman. Significance levels are denoted as follows: * p<0.10, ** p<0.05, *** p<0.01. Sample: Rural women aged 15-49 years old.

Source: DHS Benin 2011 and 2017.

tendance do not extend to secondary school attendance, as we find no significant effect of the program on this variable. It strongly suggests that the impact that we capture is indeed resulting from the exposure to primary school building rather than an exposure to better economic conditions spurred by growth. It provides additional evidence that our identification strategy is valid.

Was men's education impacted by the policy? We now look at the impact of the policy on boys' attendance. The policy was not designed to specifically target girls, thus boys could also have been impacted. The most salient point is that the reform did not seem to have an impact on boys' education. We explain primary school attendance using the same double-difference strategy as for women. Table 1.6 displays the results indicating that boys are not significantly impacted by the reform. This absence of change in the trend of boys' education can be explained by their already greater access to schooling before (and even after) the reform. A 2002 World Bank report[30] estimated a 22 percentage points difference in access to primary school between boys and girls in rural areas (86% for boys versus 64% for girls).

Because boys' primary school attendance was not affected by the school construction program, we can rule out boys' education as a potential confounder of the effect that we may capture on fertility outcomes of women and their tolerance of IPV.[31]

More specifically, the potential confounder we would like to rule out is an increase in the education of the men susceptible to becoming the husbands of the women in our sample.[32] Specific factors in this context undermine the assumption that husbands alone could drive the potential effects on age at marriage and tolerance of IPV. For example, the mean difference in age between partners in Benin is 8 years. This means that the average husband was not impacted by the reform because he was too old to have benefited from it, with the youngest women in our sample being a possible exception. Yet, only 5% of the husbands of treated women were born in 1989 or after. Table A-1.7 in the Appendix provides evidence that the husbands of women in our study cohorts were not impacted by the reform. This absence of any impact on boys' and husbands' primary school attendance means that any potential benefits to various

[30] World Bank Country Status Report: "'The Beninese education system, performance and room for improvement for the education policy"', 2002.

[31] One may argue that age at marriage and IPV tolerance are not affected (only) by a woman's own education but also by her partner's. We may overestimate the effects of the policy on those variables going through women's own outcomes, if the husbands of the women in our sample are also impacted by the reform.

[32] Here, we refer to the risk that potential effects on age at marriage and on tolerance of IPV could be originally linked to an increase in men's but not women's education. Note that we do not exclude the possibility that an increase in the husband's level of education could also explain the results as a channel and not a confounding effect. However, in this case, the increase in the husbands' education results from the increase in women's education. We will look at this potential channel in the last section of the paper.

Table 1.6: Probability of primary school attendance for men in rural areas

	Primary school attendance
Panel A: Interest Experiment: Individuals aged 4 to 8 or 12 to 17 in 1997	
Number of school built between 1997 and 2003 * Treat	0.028
	(0.03)
Controls Individual	Yes
Mean Dep. Var.	0.53
N	1922.00
r2	0.29
F	5.82
Panel B: Placebo Experiment: Individuals aged 13 to 17 or 21 to 26 in 1997	
Number of school built between 1997 and 2003 * Placebo	-0.029
	(0.03)
Controls Individual	Yes
Mean Dep. Var.	0.40
N	1637.00
r2	0.20
F	3.06

Note: The dependent variable is having attended primary school. All specifications include municipality dummies, year of birth dummies and interactions between the year of birth dummy and the number of children in the district of birth in 1993. We control also for ethnicity, religion, age and age squared. Significance levels are denoted as follows: * p<0.10, ** p<0.05, *** p<0.01. Sample: Rural men aged 15-49 years old.

Source: DHS Benin 2011 and 2017.

marital life aspects occur as a result of the human capital acquired by girls in primary school.

Regression Kink Design

The RKD strategy yields results that are consistent with those found using the double difference strategy, as shown in the last panel of Table 1.4. For rural areas, a back of the envelope calculation based on the 8-year gap between the kink and the median of the treatment group shows that the RKD strategy yields similar estimates as those from using a double difference for attending primary school (respectively $8 * 0.006 = 0.048$, compared to 0.041). Interestingly, although the difference-in-difference estimation captures no effect when pooling rural and urban cohorts together, the RKD-inspired yields significant results. Two factors underlined earlier are probably responsible for this. First, our RKD-inspired analysis relies on cohorts born between 1974 and 1994 while the difference-in-difference cohorts are more restricted. The RKD provides more statistical power and estimates the effects over a longer period of time. Second, the RKD-inspired specification includes a year-of-birth trend rather than a fixed effect, which allows us to more exhaustively use the variation in the data and provide more precise estimates.

1.3.2 Taking into account migration

We use the number of schools built in the respondent's place of residence rather than the number of schools built in her place of birth when she was of schooling age. As the DHS data do not include information on the respondent's municipality of birth nor on whether women migrated in both DHS waves, we cannot explore the potential impacts of migration for the whole sample. We nevertheless provide elements to show that migration does not jeopardize our results.

Using the number of schools built in the respondent's place of residence rather than the number of schools built in her place of birth when she was of schooling age may raise an issue of measurement error that could weaken our instrument. But as we find a positive impact of the number of schools built on education, it suggests that the approximation does not prevent us from identifying an effect. Nevertheless, we bring different elements to show that migration is not so much a threat.

First, we use the Beninese census of 2013 and compute statistics on migration and find that 18% of women born between 1989 and 1993 have migrated between municipalities.[33] We also find that there are more migrant women who currently live in urban areas than in rural areas: according to the census, only 16.8% of women who actually reside in rural areas are migrants, while this rate is as high as 30.9% in urban areas. Since our main analysis concerns women living in rural areas, using a woman's place of residence is unlikely to jeopardize our identification strategy. And migrant women currently living in urban areas are much more educated than non-migrant women who are currently living in rural areas (71.9% have some primary education vs 32.7%). It hints at the fact that rural-born women who migrate to urban areas are on average more educated than those who remain in rural areas. It would therefore bias our estimates downward, if at all.

Another element to take into account is that, in a virilocal context, women tend to migrate to the place of their partner. One caveat would be therefore that we are capturing the impact of a school constructions in the childhood place of residence of the husband rather than in her place of residence. In this case, we would expect to see an impact on the education of men, which is not the case. We already rule out this possibility. [34]

[33] Ideally, we would have liked to match the location of schools with the women's places of birth and check whether our results on primary school attendance are consistent with the census data. However, Beninese census data are not precisely geolocalised and we are not able to perform this test. We are nevertheless aware that doing so would introduce another measurement error as women did not necessarily attend primary school in the municipality where they were born. Furthermore, both sets of data are not comparable. For instance, the share of women attending primary school is 10 percentage points lower in the census compared to DHS.

[34] We could also imagine that the schools constructions did not impact men's education, but impacted norms, and

Second, we use additional data on migration collected in the 2017-18 DHS survey wave. In this survey wave, the respondent is asked how many years she has been living in her current residence. We use this information and rerun the analysis only on women who have never migrated or who arrived in their current location before the age of eight, and therefore for whom there is no approximation in measuring the intensity of the treatment.

On this sample, results are more salient, as seen in Table A-1.8 in Appendix. The coefficient of interest when investigating the probability to attend primary school is multiplied by 3, which is consistent with the idea that the treatment is more accurately measured on this sample. Despite the fact that women who did not migrate are likely to be negatively selected[35], zooming in on this subsample nevertheless suggests that using the place of current residence rather than the place of birth of respondents tends, in this case, to underestimate the actual impact of the program on primary school attendance.

Finally, we check whether migration systematically correlates with the expansion of primary schools by identifying variables that predict migration, and by interacting them with birth cohorts in our specification if necessary. Table A-1.9 shows that the expansion of primary schools does not predict migration, unlike belonging to the ethno-linguistic groups Fon or Yoruba rather than Adja.[36] Table A-1.10 in the Appendix shows that the results on primary school attendance are nevertheless robust to interacting the ethno-linguistic groups Fon, Yoruba and Adja with the birth cohorts of the respondents.

1.3.3 Marital Outcomes

Next, we investigated changes in age at first marriage, in the probability of being married as a child and in age at first birth.[37]

Table A-1.11 in the Appendix shows that, the difference-in-difference model estimated with

therefore men's tolerance to domestic violence (which is unlikely). We test this alternative scenario, by looking at the direct impact of schools constructions on men's tolerance to domestic violence: results are insignificant. Results available upon request.

[35]Descriptive data computed from the 2013 census suggest that more educated women tend to leave their rural area of birth. Those who stay are likely to be the ones that did not manage to take advantage of their additional education to look for other opportunities outside of their village of birth.

[36]The group other nationalities represent less than 1% of our sample, so we choose to set it aside.

[37]The demographic literature has expressed concern about measurement errors for age at first marriage due to recall issues. The DHS's interviewer manual (ICF (2017)) mandates that age at first union be collected by asking women the month and year when they began cohabiting with a partner for the first time. If they do not know the year, the interviewer must probe for the year of first cohabitation and is advised to do so based on the year of the first birth collected earlier in the survey and by asking how long after the beginning of the union did the respondent give birth to her first child. If the interviewer is unable to obtain an answer for the year of first cohabitation, he or she asks the woman at which age she began cohabiting with a man for the first time. As with age at the time of the survey, if the interviewer cannot obtain an answer, he or she probes for the age by following the same procedure.

OLS finds no effect from the education policy on marital outcomes or on entry into motherhood. Yet, as mentioned earlier, this model does not take into account the right-censored nature of the data, unlike duration models. The two first panels of Table 1.7 present the results from the difference-in-difference using the discrete-time duration model and provides evidence that the education program delayed the treated cohort's entry into marital life and motherhood in rural areas. Taking age at marriage for instance, building one school per 1000 children multiplies the odds of experiencing the event by 0.88 relative to the older cohort (meaning the odds decrease). Interpreting the second stage, having a primary education multiplies the odds of experiencing the event by 0.05.[38]

We also estimate the RKD model (last panel of Table 1.7). Both the RKD and the difference-in-difference deliver significant results in marital outcomes, as shown in the last panel of Table 1.7. This is also true when we compare the magnitude of the results ($0.97^8 = 0.78$ for RKD compared to 0.882 with the double difference).[39] [40]

As earlier, we rerun the analysis only on women who have never migrated or who arrived in their current location before the age of eight. Indeed, in a context where women go and live in their husband's place of residence upon marriage, our estimation may capture the impact of women's migration on marital outcomes rather than a change in their exposure to primary school. We find that the results on age at marriage are consistent (albeit not significant, see Table A-1.12 in Appendix) with our main results, and the estimates yielded by the difference in difference for the arrival of the first child remain significant. The loss of significance for age at first marriage could be due to the fact that women who do not migrate are negatively selected; despite being more educated, they may be the one who did not manage to take advantage of their additional education to make a more advantageous match, outside of their village of birth.

1.3.4 Tolerance of IPV

Next, we shed light on changes in women's attitude regarding tolerance of IPV. Panel A in Table 1.8 provides evidence that, on average, one school built for 1000 children significantly decreases the probability of condoning wife beating for going out without telling the husband, neglecting the children, refusing sex and burning the food by roughly 2 percentage points. The effects are significant at the 10% level for burning the food and 5% for the other three men-

[38] We find similar results using a Cox model (table available upon request).
[39] Using a Cox model, results are similar (table available upon request).
[40] Using an OLS estimation, we do not find any significant impact as with the double-difference (table available upon request).

Table 1.7: Marital Outcomes - Discrete Time Duration Model

	First union	2SLS First union	First child	2SLS First child
Panel A: Interest Experiment: Individuals aged 4 to 8 or 12 to 17 in 1997				
Number of school built between 1997 and 2003 * Treat	0.882***		0.879***	
	(0.03)		(0.03)	
Primary educ		0.054***		0.051***
		(0.05)		(0.04)
Controls Individual	Yes	Yes	Yes	Yes
Mean Dep. Var.	0.05	0.05	0.05	0.05
N	114545	114545	120594	120594
r2_p	0.26	0.26	0.30	0.30
chi2	6072.17	6072.17	5861.35	5861.35
Panel B: Placebo Experiment: Individuals aged 13 to 17 21 to 26 in 1997				
Number of school built between 1997 and 2003 * Placebo	1.031		1.036	
	(0.04)		(0.04)	
Primary educ		0.058		0.025
		(0.20)		(0.10)
Controls Individual	Yes	Yes	Yes	Yes
Mean Dep. Var.	0.05	0.05	0.05	0.05
N	97662	97662	102571	102571
r2_p	0.24	0.24	0.27	0.27
chi2	5670.94	5670.94	5931.35	5931.35
Regression Kink Design				
Normalized birth year* post kink* Number of schools built	0.977***		0.983***	
	(0.00)		(0.00)	
Primary educ		0.033***		0.081***
		(0.02)		(0.05)
Controls Individual	Yes	Yes	Yes	Yes
Mean Dep. Var.	0.05	0.05	0.05	0.05
N	206146	206146	217190	217190
r2_p	0.26	0.26	0.29	0.29
chi2	10935.44	10935.44	10743.08	10743.08

Note: The dependent variable is in the following order: time before marriage, time before first child and time between first and second child. The table presents the log odd-ratio. The number of observations changes between the different outcomes, since not all women have faced such events at the time of survey. The number of observations is also higher than in the OLS estimates, because data are reshaped: one observation corresponds to one year for woman. The panel A and panel B present the results of the double difference model. All specifications include municipality dummies, year of birth dummies and interactions between the year of birth dummies and the number of children in the municipality of residence in 1993. We control also for ethnicity, religion, age, age squared and for the interaction between year of birth dummies and the attendance rate in 1993 in the municipality of residence. Significance levels are denoted as follows: * p<0.10, ** p<0.05, *** p<0.01. Sample: Rural women aged 15-49.

Source: DHS Benin 2011 and 2017.

tioned motives. Furthermore, the results are robust to multiple hypothesis testing correction using the Simes method.[41] The estimates of the impact of school constructions on the last two motives represent nearly one third of the baseline level for tolerance of IPV. The IPV index we built using a PCA conveys the same message as the item-per-item analysis. Even though we cannot completely rule out that the schooling program had effects through channels other than women's own education, the results suggest that an increase in primary education reduces the tolerance index of IPV by 1.59 standard deviation.

The placebo test of Panel B in Table 1.8 shows an increasing trend (so opposite to our effect) for some of the individual IPV items. However, those estimates of the individual items of IPV are not robust to MHT and the placebo regressions of the index of tolerance of IPV in Panel B is not significant either.

To further make our case that there is no pre-existing trend to tolerance of IPV, we split geographical areas according to the intensity of the treatment as we did earlier and plot the tolerance of IPV over birth cohorts in the two groups. As can be seen in Figure 1.7, the index of tolerance of IPV before the surge in school constructions tends to evolve in parallel in the two groups. In addition, we run regressions using the binary measure of the intensity of the construction of schools and the placebo is not significant.[42]. We also go further back in time and compare women aged 18-24 years to women aged 26-31 in 1997 and detect no trace of a previous trend in tolerance of IPV among them either.[43]

Finally, all the placebo regressions in Panel B of tables 1.4 and 1.7, are not significant which is not consistent with the existence of an endogeneous placement of schools.

The effect of the treatment on each cohort displayed in Figure 1.9 illustrates that a change in trend occurs for the younger cohorts.[44]

[41] The Simes method is a modification of the Bonferroni procedure and is based on the ordered p-values of the individual tests. See Simes (1986).

[42] Tables available upon request.

[43] Table available upon request.

[44] despite the noise probably related to age heaping that shows in the jagged evolution of the effect per age.

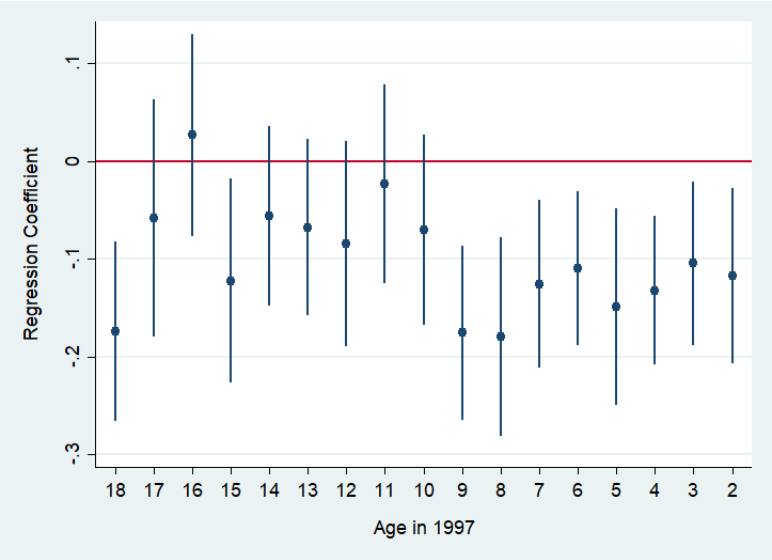
Table 1.8: Tolerance of Intimate Partner Violence

	Goes out without telling husband	Neglects the children	Argues with husband	Refuses sex	Burns the food	IPV index	Second stage IPV index
Panel A: Interest Experiment: Individuals aged 4 to 8 or 12 to 17 in 1997							
Number of school built between 1997 and 2003 * Treat	-0.023**	-0.030**	-0.010	-0.020**	-0.014*	-0.066**	
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)	
Primary educ							-1.599**
							(0.69)
Controls Individual	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Robust to MHT	Yes††	Yes††	No	Yes††	Yes†	NA	
Mean Dep. Var.	0.17	0.19	0.18	0.11	0.12	0.01	0.01
N	6241	6241	6241	6241	6241	6241	6241
r2	0.16	0.13	0.11	0.08	0.08	0.14	-0.15
F	4.05	5.70	3.09	2.69	2.98	4.62	13.84
Panel B: Placebo Experiment: Individuals aged 13 to 17 or 21 to 26 in 1997							
Number of school built between 1997 and 2003 * Placebo	0.007	0.020*	0.008	0.016*	0.001	0.035	
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)	
Primary educ							-6.077
							(14.12)
Controls Individual	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Robust to MHT	No	No	No	No	No	NA	
Mean Dep. Var.	0.15	0.17	0.17	0.10	0.11	-0.04	-0.04
N	5117	5117	5117	5117	5117	5117	5117
r2	0.14	0.12	0.10	0.07	0.09	0.12	-3.70
F	4.53	5.44	3.10	2.37	3.17	4.59	2.65
Regression Kink Design							
Normalized birth year* post kink*	-0.00062	-0.00214*	-0.00020	-0.00100	-0.00190**	-0.00521*	
Number of schools built							
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	
Primary educ							-0.83061*
							(0.49)
Controls Individual	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Robust to MHT	Yes‡	No	No	No	Yes‡	NA	
Mean Dep. Var.	0.16	0.18	0.18	0.11	0.11	-0.01	-0.01
N	11178	11178	11178	11178	11178	11178	11178
r2	0.14	0.11	0.09	0.06	0.07	0.12	0.06
F	7.00	10.11	5.14	4.55	5.76	8.29	46.16

Note: The dependent variable is in the following order a dummy taking the value 1 if the woman finds wife beating acceptable if a woman goes out without telling her partner, argues with him, neglects the children, refuses sex and burns the food. We control also for ethnicity, religion, age and age squared. Panel A and panel B present the results of the double difference model. Significance levels are denoted as follows: * p<0.10, ** p<0.05, *** p<0.01. Sample: Rural women aged 15-49 years old.

Source: DHS Benin 2011 and 2017.

Figure 1.9: Effect of the treatment on the IPV index



Note: The figure presents the coefficients of the interaction of respondent's age in 1997 and the number of schools built between 1997 and 2003 in the region of residence in equation (2). The dependent variable is an index built with a PCA. Sample: Rural women aged 15-49 years old.
Source: DHS Benin 2011 and 2017.

We may want to disentangle what a change in tolerance of IPV over time means. As mentioned in Section 2.2, the DHS questions that assess tolerance of IPV rely on two perceptions: first, that the behavior in each scenario transgresses gender norms; and, second, that a woman believes it falls on the husband to chastise her using violence. The effect of primary school that we capture may indicate that: (1) education modified gender norms and relaxed the expectation around women's behavior in the household; (2) increased education changed women's perceptions of husbands' alleged right to use physical violence to police their behavior; or (3) that both phenomena occurred simultaneously. The first would suggest that the education policy relaxed the constraint on women's expected behavior in their households. As a result, this would mean that access to schooling can foster a process of awareness around the gender norms that curtail a woman's ability to choose, for instance, when to have intercourse with their partner. If access to primary school impacted only women's opinions of using violence to police wives' behavior, this would also challenge another norm, i.e. husband's right to use violence to chastise his wife. Either way, both pathways are synonymous with improving women's empowerment and well-being.

We may again be concerned that the results are driven by women who are married to men whose education has been impacted by the reform, if not through their education, through the exposure to an increasingly gender-mixed environment. As shown earlier, the policy studied does not seem to have a significant impact on men's education, and only 5% of the husbands

of treated women were born in 1989 or after. Still, we conducted an analysis on the tolerance to domestic violence, this time removing from the sample women who married husbands potentially affected by the reform (Table A-1.13 in the Appendix). The results on tolerance to domestic violence remain unchanged. Let it be clear that we are not claiming that matching has no part at all in the identified effect. It is plausible that women who are more educated as a result of the policy tend to wed more educated husbands. In this case, the husband's education's in particular serves as a channel when it comes to domestic violence, but it is a channel activated by the increased education of women.

Using an RKD-inspired method, we find that the estimates (displayed in the last panel of the Table 1.8) are consistent with those found using the double-difference strategy: the magnitude is in general lower (with the coefficient when the index is the outcome being, respectively, $-8 * 0.00953 = -0.08$, and -0.12). For the items "Goes out without telling the husband" and "Refuses sex", the results turn out to be insignificant, although the direction of the effect remains consistent with the double difference estimates. As outlined in Section 1.2, the differences between the two strategies' estimates are likely to be partially driven by more flexible control of the time effects in the RKD-inspired specification. In addition, the RKD exploits more of the variation coming from the linearity of the treatment than does the difference in difference estimation.

Redoing the analysis only on women who have never migrated or who arrived in their current location before the age of eight, we do not find a significant impact of schools construction on measures of tolerance of IPV[45], which could again be due to the fact that women who do not migrate, though more educated, are likely to be negatively selected. It is nevertheless worth noting that, using a binary version of the treatment (with a dummy taking the value 1 if the number of schools is above the 75th percentile), results are significant with the difference in difference for all items except "burns the food" (Table A-1.14). Results are also significant when using the aggregated index of tolerance of IPV.

We also explore whether the education reform has an impact on women's say in the decision-making in the household, which is a common measure of women's agency. Table A-1.16 shows that the construction of schools does not impact women's reported involvement in decisions in the household suggesting that all dimension of women's empowerment were not equally impacted by the program.

[45]Table available upon request

Table 1.9: Experience of IPV

	score control	emotional ipv	physical ipv
Panel A: Individuals aged 4 to 8 or 12 to 17 in 1997			
Number of school built between 1997 and 2003 * Treat	-0.046	-0.064*	0.020
	(0.14)	(0.04)	(0.05)
Controls Individual	Yes	Yes	Yes
Mean Dep. Var.	0.02	0.40	0.24
N	1013	1013	1013
r2	0.24	0.22	0.16
F	2.21	3.19	3.45
Panel B: Placebo Experiment: Individuals aged 13 to 17 or 21 to 26 in 1997			
Number of school built between 1997 and 2003 * Placebo	0.147	0.064	-0.058
	(0.17)	(0.09)	(0.10)
Controls Individual	Yes	Yes	Yes
Mean Dep. Var.	-0.08	0.40	0.24
N	662	662	662
r2	0.32	0.29	0.29
F	2.94	2.17	2.63

Note: The dependent variables the standardized score of controling behavior, a dummy variable measuring emotional violence, and a dummy variable measuring physical ipv or marital rape. All specifications include municipality dummies, year of birth dummies and interactions between the year of birth dummy and the number of children in the district of birth in 1993. We control also for ethnicity, religion, age and age squared. Significance levels are denoted as follows: * p<0.10, ** p<0.05, *** p<0.01.

Source: DHS Benin 2017. Women residing in rural areas. Sample: subsample of ever-partnered women selected to answer the DHS domestic violence module. One eligible woman per household is selected.

1.3.5 Experience of IPV

Using the subsample of respondents who answered the DHS module on domestic violence in the DHS 2017-18 survey wave, we find that the education reform decreased by 6.4pp the experience of emotional violence, which represents a decrease of 16% compared to the level of IPV in the control group. The result is significant at the 10% level. This is consistent with the idea that the atmosphere within the couple could be less threatening to women who were impacted by the program. We do not find an impact on the score of controling behavior nor on the experience of physical IPV, though.

1.4 Channels explaining the decrease in tolerance to IPV

We may expect that one key explanation for the results on tolerance of IPV depends on the husband's and wife's relationship. We use a series of robustness tests to check that husbands or potential husbands had not been affected by the reform. Therefore, we can interpret our results as a consequence of an increase in women's education rather than of improved men's education. Men's education serves as a pure channel and not a confounding effect. Yet, we cannot sufficiently conclude that men's pathway do not play a part in the chain of mechanisms

leading to our results on women's well-being. Indeed, even though men are not affected on average, more educated women potentially have access to more educated men in the marriage market, thus making for a better match. This could explain what we see in terms of domestic violence. Ideally, we would have liked to have data on all potential (unrealized) matches. Nevertheless, our data allows us to look at some characteristics of the realized matches for married women at the time of the survey. However, it is worth noting that, the reform being relatively recent, treated women were not all married at the time of the survey. 20.3% of the treated women had never been in a union at the time of the survey, whereas this is the case for only 1.45% of the women in the control group. As mentioned earlier, women who were already married are likely to be selected in a particular way, as well as their husbands. Our analysis on the characteristics of the husbands is therefore likely to be biased.

Another point to keep in mind is that we conducted our main analysis on both married women and unmarried women. However, tolerance of IPV is unlikely to have the same meaning for women who never lived with a partner and for those who have an experience of marital life. We first check whether the results on IPV are driven by married or unmarried women, then looked at the husbands' characteristics. We investigate the results according to the marital status in the Table 1.10 and found that the change in tolerance of IPV is driven by married women. It is important to keep in mind that, among the treated groups, married women are likely to exhibit some vulnerability compared to their unmarried counterparts. This could mean our results are biased downwards. Alternatively, we may believe that there is more room for improvement among those more vulnerable women: in this case, the direction of the bias would be reversed. Thus, the overall effect is unclear. This feature urges us to be modest in our interpretation of the education policy's impact on tolerance of IPV.[46]

In looking at the age gap with the partner, the sign of the coefficient of interest is negative but not significant for all married women (Table A-1.15 in the Appendix). This hints at the fact that, in our specific case, improvements in women's attitudes regarding tolerance of physical abuse is not driven by a change in their partner's age profile. Next, we look at the difference in education relative to the husband (Table A-1.15) and find that the education policy did not decrease the education gap with the partner. This results partially from the fact that not every woman in the treatment group was married at the time of the survey, as underlined earlier, leads to issues related to selecting particuliar women from among the married.

It stems from this analysis that the impact of the program on the tolerance of IPV is not driven

[46]It would be useful to look at the same impact with posterior data, in order to see whether the effect holds when every woman got married.

Table 1.10: Interaction with marital status

	Goes out without telling husband	Neglects the children	Argues with husband	Refuses sex	Burns the food	IPV index
Panel A: Interest Experiment: Individuals aged 4 to 8 or 12 to 17 in 1997						
Number of school built between 1997 and 2003 * Treat	0.018	0.009	0.043*	-0.004	-0.007	0.072
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.13)
Number of school built between 1997 and 2003* Ever married*Treat	-0.048**	-0.045**	-0.054**	-0.018	-0.007	-0.213
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.13)
Controls Individual	Yes	Yes	Yes	Yes	Yes	Yes
Robust to MHT	Yes†	Yes†	Yes†	No	No	NA
Mean Dep. Var.	0.17	0.19	0.18	0.11	0.12	0.02
N	6241	6241	6241	6241	6241	6241
r2	0.16	0.13	0.12	0.08	0.08	0.14
F	4.04	5.81	3.27	2.93	3.06	4.83

Note: The dependent variable is in the following order a dummy taking the value 1 if the woman finds wife beating acceptable if a woman goes out without telling her partner, argues with him, neglects the children, refuses sex and burns the food. All specifications include municipality dummies, year of birth dummies and interactions between the year of birth dummy and the number of children in the district of birth in 1993. We control also for ethnicity, religion, age and age squared. Significance levels are denoted as follows: * p<0.10, ** p<0.05, *** p<0.01. Sample: Married rural women aged between 15 and 49.

Source: DHS Benin 2011 and 2017.

by a change in the relative characteristics of women compared to their partners. It may therefore be linked to changes pertaining to women themselves. We explore therefore the potential returns on the labor market women may have benefited from as a result from the education policy. We only present reduced form estimates. In rural areas, 71.15% of women were currently working at the time of the survey. Among women who worked at the time of the survey, the main occupations were working in sales (31.44%), working in the agricultural sector (27.68%), and working in the services sector (18.98%). The services sector includes working in hospitality for instance. Table 1.11 shows that the education reform appears to have decreased women's probability to be working at the time of the survey by 5% compared to the baseline level. Among women who worked at the time of survey, treated women are less likely to work in sales (-22% compared to the mean in the control group) and more likely to work in the services sector (+12%). The DHS data does not allow us to explore the reasons why respondents exposed to the education reform are less likely to be working at the time of the survey. Treated women might be less likely to work because they may have married into more wealthy families, and as a result do not have to work (Cameron *et al.* (2001)). It is difficult to test in practice though as the level of wealth of the household is likely to be endogenous to the education of the respondent. Another hypothesis might be that the education reform increased the reservation wage of treated women, which makes them more likely to be unemployed at the time of the survey. The results on the decrease of the tolerance to violence seems therefore not to be linked with a higher bargaining power allowed by a higher financial autonomy, even if we have to keep in mind that the sole increase in the ability to work (and of potential returns of education on the labor market) can increase the bargaining power of women, even if they do not actually choose to work.

Table 1.11: Respondent's literacy and occupation

	literacy	currently working	works in agriculture	works in sales	works in services
Panel A: Individuals aged 4 to 8 or 12 to 17 in 1997					
Number of school built between 1997 and 2003 * Treat	0.019	-0.039***	0.015	-0.067***	0.023*
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
Controls Individual	Yes	Yes	Yes	Yes	Yes
Mean Dep. Var.	0.20	0.74	0.33	0.30	0.19
N	6221	6241	4692	4692	4692
r2	0.16	0.17	0.24	0.11	0.09
F	10.31	10.68	4.56	4.30	3.05
Panel B: Placebo Experiment: Individuals aged 13 to 17 or 21 to 26 in 1997					
Number of school built between 1997 and 2003 * Placebo	0.007	0.008	0.007	-0.001	0.013
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
Controls Individual	Yes	Yes	Yes	Yes	Yes
Mean Dep. Var.	0.10	0.80	0.36	0.31	0.19
N	5103	5117	4150	4150	4150
r2	0.08	0.20	0.24	0.14	0.10
F	2.31	5.77	3.63	2.58	3.40

Note: Women residing in rural areas. The dependent variable in column 1 is a binary variable equal to 1 if women were able to read parts or a whole sentence on a card presented by the DHS enumerator. The dependent variable in column 2 is a binary variable equal to 1 if the respondent was working at the time of the survey. The dependent variable in column 1 is a binary variable equal to 1. The dependent variables in column 3, 4 and 5 are binary variables equal to 1 if the respondent works in agriculture, in sales or in services. All specifications include municipality dummies, year of birth dummies and interactions between the year of birth dummy and the number of children in the district of birth in 1993. We control also for ethnicity, religion, age and age squared. Significance levels are denoted as follows: * p<0.10, ** p<0.05, *** p<0.01.

Source: DHS Benin 2010 and 2017. Women residing in rural areas.

We therefore explore another potential channel to explain the decrease in the tolerance to IPV: the impact of the program on what has been learned at school. We know that the curricula of the 90's were not focused on the place of women in the society, or on the rights of women: therefore it is probably not the main channel. We also know that the share of female teachers at primary school was really low (20% according to the World Bank data), so a mechanism of "role model" seems also not to be at stake. We explore the impact of the program on a simple outcome, literacy. A the time of the survey, in rural areas, the level of literacy is low with 29.45% of women being able to read a parts or a whole sentence written on a card presented by the DHS enumerator. As shown in Table 1.11, the education reform does not seem to have had a significant impact on women's literacy, which is consistent with a context where the quality of education is low; as a result, increasing enrolment is unlikely to improve significantly pupils' schooling performance. We note that the results are nevertheless marginally significant (at 15%). One possible hypothesis explaining the results on the tolerance to IPV could be that attending primary school improved women's self-esteem, either because they got better at mastering writing and reading than the previous generation of women in their household, or because of the mere fact of attending school, alongside boys who were till then more likely to be sent to school, even if the curriculum was not focusing on women's rights.

For a subsample of married women, we were able to analyze the husbands' tolerance of IPV and found that the policy decreased their tendency to condone domestic violence, as measured by the IPV index shown in Table 1.12. This is interesting, since these men were not directly treated by the policy (having been too old). This means that either the policy has created a

Table 1.12: Tolerance of IPV of husbands

	Goes out without telling husband	Neglects the children	Argues with husband	Refuses sex	Burns the food	IPV index
Panel A: Interest Experiment: Individuals aged 4 to 8 or 12 to 17 in 1997						
Number of school built between 1997 and 2003 * Treat	-0.008	-0.011	-0.038	-0.032**	-0.024*	-0.195*
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.11)
Controls Individual	Yes	Yes	Yes	Yes	Yes	Yes
Robust to MHT	No	No	No	No	No	NA
Mean Dep. Var.	0.08	0.08	0.10	0.07	0.06	0.03
N	1712	1712	1712	1712	1712	1712
r2	0.15	0.14	0.20	0.12	0.18	0.19
F	1.05	1.21	1.68	0.91	1.41	1.07
Panel B: Placebo Experiment: Individuals aged 13 to 17 or 21 to 26 in 1997						
Number of school built between 1997 and 2003 * Placebo	0.013	0.016	0.042**	0.019	0.003	0.155*
	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.08)
Controls Individual	Yes	Yes	Yes	Yes	Yes	Yes
Robust to MHT	No	No	No	No	No	NA
Mean Dep. Var.	0.08	0.07	0.10	0.06	0.05	-0.04
N	1422	1422	1422	1422	1422	1422
r2	0.18	0.14	0.20	0.14	0.15	0.19
F	0.97	1.07	1.67	0.86	0.88	1.39

Note: The dependent variable is in the following order a dummy taking the value 1 if the man finds wife beating acceptable if a woman goes out without telling her partner, argues with him, neglects the children, refuses sex and burns the food. All specifications include municipality dummies, year of birth dummies and interactions between the year of birth dummies and the number of children in the municipality of residence in 1993 for the wife. We control also for ethnicity, religion, age, age squared and for the interaction between year of birth dummies and the attendance rate in 1993 in the municipality of residence of the wife. Significance levels are denoted as follows: * p<0.10, ** p<0.05, *** p<0.01. Sample: men married with a sub-sample of women aged between 15 and 49. Correction for MHT is not applied to the IPV index.

Source: DHS Benin 2011 and 2017.

more general change in gender norms, or that they could have been influenced by their wives.

To summarize, the main drivers of our results are likely to be, on the one hand, improvements in women's (and their husbands') own opinions regarding how they should be treated and, on the other hand, their increased bargaining power due to schooling.

1.5 Conclusion

Using a double-difference strategy along with a method inspired by a regression kink design, we show that a 1990s rise in school constructions in Benin, designed to reach the Millenium Development Goals, increased the primary school attendance of women living in rural areas. We find evidence that the policy intervention improved women's marital outcomes by decreasing the tolerance of wife-beating and by delaying entry into marital life and motherhood. We also investigate the pathways through which the policy impacts women's well-being, and our results hint at the fact that the effect is not driven by an evolving profile of women's partners in terms of age difference or education, but rather by a change in women's own outcomes. This education policy, which targets essentially the supply side of education, appears to have been successful beyond its initial agenda, namely by impacting crucial aspects of women's well-being. These are the first results that zero in on the effects of 1990s mass education programs led in the nineties in West Africa, specifically on women's well-being within the household. Other West African countries have experienced similar policies, and further research on their impacts may help build a comprehensive understanding of their efficiency.

Appendix

Table A-1.1: School Construction by District since the 1980's

	Region level				
	Stock 1979	*1980-1989*	*1990-1999*	*2000-2005*	*Stock 2005*
Mean	175.17	53.42	89.42	113.25	431.17
min	99	23	56	27	299
max	284	84	180	170	562
median	162	54	84.5	125	454.5
N	12	12	12	12	12

Source: PASEC data on school constructions in Benin.

Table A-1.2: Correlation with outcomes before the program

Variables	Urban			Rural		
	High Treatment	Low Treatment	Diff.	High Treatment	Low Treatment	Diff.
Marriage before 15	0.10	0.15	0.05***	0.16	0.18	0.02
			(0.01)			(0.28)
Index tolerance to IPV	-0.15	-0.07	0.08	-0.07	0.02	0.09
			(0.75)			(0.34)
Goes out without telling the husband	0.10	0.13	0.03	0.16	0.16	0.00
			(0.10)			(0.87)
Neglects the children	0.11	0.13	0.03	0.17	0.18	0.01
			(0.21)			(0.44)
Argues with the husband	0.10	0.14	0.04*	0.16	0.18	0.02
			(0.06)			(0.27)
Refuses sex	0.06	0.09	0.03*	0.10	0.12	0.02
			(0.07)			(0.25)
Burns the food	0.06	0.10	0.04**	0.11	0.12	0.02
			(0.01)			(0.31)
Number of clusters	201	206	407	354	358	712

Note: The first column (respectively the fourth) displays the the mean of the outcomes by cluster before the program, in areas where the number of school built for 1000 children between 1997 and 2003 is higher than the median, in urban areas (respectively in rural areas). The second column (respectively the fifth) displays the mean of the outcomes by cluster before the program, in areas where the number of school built between 1997 and 2003 is lower than the median, in urban areas (respectively in rural areas). The third and the sixth columns represent results of the ttest. Significance levels are denoted as follows: * p<0.10, ** p<0.05, *** p<0.01. Sample: DHS clusters. Each mean is computed on women who were between 13 and 17 years old (not treated) in 1997. Cotonou is excluded.

Source: DHS Benin 2011 and 2017.

Table A-1.3: Smoothness of Covariates

	Fon	Traditional	Muslim	Christian
Normalized birth year* post kink* Number of schools built	0.001	0.000	0.001	0.000
	(0.00)	(0.00)	(0.00)	(0.00)
Mean Dep. Var.	0.38	0.15	0.27	0.51
N	11178	11178	11178	11178
r2	0.71	0.22	0.57	0.31
F	2.22	4.56	0.46	5.17

Note: The dependent variable is the number of observations by cohorts. Models (1) represents the simple regression kink design. Models (2) includes also a dummy indicating whether the cohort is younger than the kink. The bandwidth is 10 years (on both sides of the kink). Significance levels are denoted as follows: * p<0.10, ** p<0.05, *** p<0.01. Sample: Eligible women aged 15-49 years old.

Source: DHS Benin 2011 and 2017.

Table A-1.4: McCrary Test for RKD Design - Benin

	Density	
	(1)	(2)
Normalized birth year* post kink* Number of schools built	-0.00	-0.00
	(0.00)	(0.00)
Number of cohorts	7384.00	7384.00
r2	0.03	0.05
F	8.99	41.94

Note: The dependent variable is the number of observations by cohorts. Models (1) represents the simple regression kink design. Models (2) includes also a dummy indicating whether the cohort is younger than the kink. The bandwidth is 10 years (on both sides of the kink). Significance levels are denoted as follows: * p<0.10, ** p<0.05, *** p<0.01. Sample: Eligible women aged 15-49 years old.

Source: DHS Benin 2011 and 2017.

Table A-1.5: Probability of primary school attendance : variation of the treated cohorts

	1980-1985 and 1989-1993	1975-1985 and 1989-1993	1980-1988 and 1989-1993	1980-1985 and 1989-1996	1975-1988 and 1989-1996
Number of school built between 1997 and 2003 * Treat	0.041***	0.038***	0.031**	0.042***	0.042***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Controls Individual	Yes	Yes	Yes	Yes	Yes
Mean Dep. Var.	0.27	0.23	0.27	0.36	0.29
N	6241	8489	7571	7600	12130
r2	0.21	0.21	0.19	0.26	0.24
F	13.93	14.03	12.82	25.39	24.08

Note: The dependent variable is having attended primary school. All specifications include municipality dummies, year of birth dummies and interactions between the year of birth dummies and the number of children in the municipality of residence in 1993. They include also controls for ethnicity, religion, age, age squared and for the interaction between year of birth dummies and the attendance rate in 1993 in the municipality of residence. Significance levels are denoted as follows: * p<0.10, ** p<0.05, *** p<0.01. Sample: Rural women aged 15-49 years old.

Source: DHS Benin 2011 and 2017.

Table A-1.6: Probability of secondary school attendance

	Secondary school attendance
Panel A: Interest Experiment: Individuals aged 4 to 8 or 12 to 17 in 1997	
Number of school built between 1997 and 2003 * Treat	0.008
	(0.01)
Controls Individual	Yes
Mean Dep. Var.	0.12
N	6241
r2	0.16
F	10.21
Panel B: Placebo Experiment: Individuals aged 13 to 17 or 21 to 26 in 1997	
Number of school built between 1997 and 2003 * Placebo	0.005
	(0.01)
Controls Individual	Yes
Mean Dep. Var.	0.04
N	5117
r2	0.07
F	1.66

Note: The dependent variable is having attended secondary school. All specifications include municipality dummies, year of birth dummies and interactions between the year of birth dummies and the number of children in the municipality of residence in 1993. They include also controls for ethnicity, religion, age, age squared and for the interaction between year of birth dummies and the attendance rate in 1993 in the municipality of residence .Significance levels are denoted as follows: * $p<0.10$, ** $p<0.05$, *** $p<0.01$. Sample: Rural women aged 15-49 years old.

Source: DHS Benin 2011 and 2017.

Table A-1.7: Probability of primary school attendance for husbands

	School attendance of the husband
Panel A: Interest Experiment: Individuals aged 4 to 8 or 12 to 17 in 1997	
Number of school built between 1997 and 2003 * Treat	-0.016
	(0.07)
Controls Individual	Yes
Mean Dep. Var.	0.81
N	5440
r2	0.11
F	3.48
Panel B: Placebo Experiment: Individuals aged 13 to 17 or 21 to 26 in 1997	
Number of school built between 1997 and 2003 * Placebo	0.107
	(0.08)
Controls Individual	Yes
Mean Dep. Var.	0.75
N	4919
r2	0.10
F	2.74

Note: The dependent variable is having a husband who has attended primary school. All specifications include municipality dummies, year of birth dummies and interactions between the year of birth dummy and the number of children in the district of birth in 1993. We control also for the ethnicity and the religion. Significance levels are denoted as follows: * p<0.10, ** p<0.05, *** p<0.01. Sample: Rural women aged 15-49 years old.

Source: DHS Benin 2011 and 2017.

Table A-1.8: Probability of primary school attendance, Non migrant Women

	All	Urban	Rural
Panel A: Interest Experiment: Individuals aged 4 to 8 or 12 to 17 in 1997			
Number of school built between 1997 and 2003 * Treat	0.055**	-0.013	0.128***
	(0.03)	(0.04)	(0.03)
Controls Individual	Yes	Yes	Yes
Mean Dep. Var.	0.28	0.36	0.23
N	2670	978	1692
r2	0.27	0.36	0.28
F	7.46	6.64	6.10
Panel B: Placebo Experiment: Individuals aged 13 to 17 or 21 to 26 in 1997			
Number of school built between 1997 and 2003 * Placebo	0.023	0.037	-0.015
	(0.03)	(0.04)	(0.03)
Controls Individual	Yes	Yes	Yes
Mean Dep. Var.	0.16	0.23	0.11
N	1678	631	1047
r2	0.23	0.39	0.15
F	2.76	2.81	1.83
Regression Kink Design			
Normalized birth year*			
post kink*	0.006**	0.000	0.015***
Number of schools built			
	(0.00)	(0.00)	(0.00)
Controls Individual	Yes	Yes	Yes
Mean Dep. Var.	0.26	0.35	0.20
N	4435	1619	2816
r2	0.26	0.32	0.25
F	20.77	12.80	14.91

Note: The dependent variable is having attended primary school. The panel A and panel B present the results of the double difference model. We control by the ethnicity and the religion of the woman. Significance levels are denoted as follows: * p<0.10, ** p<0.05, *** p<0.01. Sample: Rural women aged 15-49 years old, who have never migrated or who arrived in their current location before the age of 8 years old.

Source: DHS Benin 2017.

Table A-1.9: Predictors of immigration after 8 years old

	Immigration after 8 years old
Number of school built between 1997 and 2003	-0.057
	(0.04)
Number of school built between 1997 and 2003 * Treat	-0.022
	(0.03)
Reference: adja	
bariba	-0.156
	(0.10)
dendi	-0.109
	(0.12)
fon	-0.152**
	(0.06)
yoa	0.159
	(0.11)
betamaribe	0.116
	(0.11)
peulh	-0.076
	(0.10)
yoruba	-0.163**
	(0.08)
other bÃ©ninois	-0.053
	(0.11)
other nationalities	0.322***
	(0.11)
Reference: poorest wealth quintile	
poorer	0.019
	(0.03)
middle	0.016
	(0.03)
richer	0.051
	(0.04)
richest	0.207***
	(0.05)
Mean Dep. Var.	0.39
N	2786
r2	0.20
F	2.60

Note: In column (1), the dependent variable is having migrated to the cluster of residence at the time of the survey after 8 years old. Significance levels are denoted as follows: * p<0.10, ** p<0.05, *** p<0.01. Sample: Rural women aged 15-49 years old.

Table A-1.10: Primary Education controling for predictors of migration and school implementation interacted with birth cohorts

	Primary Education	Years of Education
Panel A: Individuals aged 4 to 8 or 12 to 17 in 1997		
Number of school built between 1997 and 2003 * Treat	0.029*	0.114
	(0.02)	(0.13)
Controls Individual	Yes	Yes
Mean Dep. Var.	0.27	1.71
N	6241	6241
r2	0.22	0.22
F	11.31	10.11
Panel B: Placebo Experiment: Individuals aged 13 to 17 or 21 to 26 in 1997		
Number of school built between 1997 and 2003 * Placebo	-0.014	-0.022
	(0.01)	(0.10)
Controls Individual	Yes	Yes
Mean Dep. Var.	0.14	0.72
N	5117	5117
r2	0.12	0.12
F	2.41	2.19

Note: In column (1), the dependent variable is having attended primary school. In column (2), the dependent variable is the years of education of a respondent. We control for ethnicity, religion, age and age squared. Being Fon, Adja or Yoruba is interacted with the birth cohort as it predicts school implementation or migration after 8 years old.Panel A and panel B present the results of the double difference model. Significance levels are denoted as follows: * $p<0.10$, ** $p<0.05$, *** $p<0.01$, † $p<0.11$. Sample: Rural women aged 15-49 years old.

Table A-1.11: Marital Outcomes - OLS

	Age at marriage	Marriage before 15	Age at first child
Panel A: Interest Experiment: Individuals aged 4 to 8 or 12 to 17 in 1997			
Number of school built between 1997 and 2003 * Treat	0.129	0.006	0.170
	(0.16)	(0.01)	(0.12)
Controls Individual	Yes	Yes	Yes
Mean Dep. Var.	18.02	0.15	19.01
N	5574	6241	5436
r2	0.09	0.05	0.11
F	6.33	3.29	9.63
Panel B: Placebo Experiment: Individuals aged 13 to 17 or 21 to 26 in 1997			
Number of school built between 1997 and 2003 * Placebo	-0.072	-0.009	-0.137
	(0.17)	(0.01)	(0.14)
Controls Individual	Yes	Yes	Yes
Mean Dep. Var.	18.87	0.17	19.80
N	5049	5117	5022
r2	0.07	0.04	0.09
F	2.50	1.43	3.16

Note: The dependent variable is in the following order: age at marriage, marriage before 15 years old, and lastly age at first child. The number of observations changes slightly between the different outcomes, since not all women have faced such events at the time of survey. All specifications include municipality dummies, year of birth dummies and interactions between the year of birth dummies and the number of children in the municipality of residence in 1993. We control also for the ethnicity, religion, age, age squared and for the interaction between year of birth dummies and the attendance rate in 1993 in the municipality of residence. Significance levels are denoted as follows: * $p<0.10$, ** $p<0.05$, *** $p<0.01$. Sample: Rural women aged 15-49.

Source: DHS Benin 2011 and 2017.

Table A-1.12: Marital Outcomes - Discrete Time Duration Model, Non migrant Women

	First union	2SLS First union	First child	2SLS First child
Panel A: Interest Experiment: Individuals aged 4 to 8 or 12 to 17 in 1997				
Number of school built between 1997 and 2003 * Treat	0.853		0.831*	
	(0.09)		(0.08)	
Primary educ		0.340		0.276*
		(0.24)		(0.19)
Controls Individual	Yes	Yes	Yes	Yes
Mean Dep. Var.	0.05	0.05	0.05	0.05
N	31615	31615	33307	33307
r2_p	0.26	0.26	0.31	0.31
chi2	1697.93	1697.93	1823.01	1823.01
Panel B: Placebo Experiment: Individuals aged 13 to 17 21 to 26 in 1997				
Number of school built between 1997 and 2003 * Placebo	1.182		1.143	
	(0.19)		(0.17)	
Primary educ		0.013		0.003
		(0.05)		(0.02)
Controls Individual	Yes	Yes	Yes	Yes
Mean Dep. Var.	0.05	0.05	0.05	0.05
N	20105	20105	20797	20797
r2_p	0.24	0.24	0.30	0.30
chi2	1217.07	1217.07	1462.73	1462.73
Regression Kink Design				
Normalized birth year* post kink* Number of schools built	0.986		0.988	
	(0.01)		(0.01)	
Primary educ		0.423		0.488
		(0.29)		(0.32)
Controls Individual	Yes	Yes	Yes	Yes
Mean Dep. Var.	0.05	0.05	0.05	0.05
N	52721	52721	55424	55424
r2_p	0.25	0.25	0.30	0.30
chi2	2786.78	2786.78	3005.21	3005.21

Note: The dependent variable is in the following order: time before marriage, time before first child and time between first and second child. The table presents the log odd-ratio. The number of observations changes between the different outcomes, since not all women have faced such events at the time of survey. The number of observations is also higher than in the OLS estimates, because data are reshaped: one observation corresponds to one year for woman. The panel A and panel B present the results of the double difference model. All specifications include municipality dummies, year of birth dummies and interactions between the year of birth dummies and the number of children in the municipality of residence in 1993. We control also for ethnicity, religion, age, age squared and for the interaction between year of birth dummies and the attendance rate in 1993 in the municipality of residence. Significance levels are denoted as follows: * p<0.10, ** p<0.05, *** p<0.01. Sample: Rural women aged 15-49, who have never migrated or who arrived in their current location before the age of 8.

Source: DHS Benin 2017.

Table A-1.13: Tolerance to IPV - Women whose husband is not affected by the policy

	Goes out without telling husband	Neglects the children	Argues with husband	Refuses sex	Burns the food
Panel A: Interest Experiment: Individuals aged 4 to 8 or 12 to 17 in 1997					
Number of school built between 1997 and 2003 * Treat	-0.025**	-0.031**	-0.011	-0.022***	-0.014*
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Controls Individual	Yes	Yes	Yes	Yes	Yes
Mean Dep. Var.	0.17	0.19	0.18	0.11	0.12
N	6095	6095	6095	6095	6095
r2	0.16	0.13	0.12	0.08	0.08
F	3.90	5.61	3.08	2.67	3.11
Panel B: Placebo Experiment: Individuals aged 13 to 17 or 21 to 26 in 1997					
Number of school built between 1997 and 2003 * Placebo	0.006	0.021*	0.008	0.017**	-0.001
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Controls Individual	Yes	Yes	Yes	Yes	Yes
Mean Dep. Var.	0.15	0.16	0.17	0.10	0.11
N	5065	5065	5065	5065	5065
r2	0.14	0.12	0.10	0.07	0.09
F	4.51	5.54	3.06	2.40	3.10

Note: The dependent variable is in the following order a dummy taking the value 1 if the woman finds wife beating acceptable if a woman goes out without telling her partner, argues with him, neglects the children, refuses sex and burns the food. All specifications include municipality dummies, year of birth dummies and interactions between the year of birth dummies and the number of children in the municipality of residence in 1993. We control also for ethnicity, religion, age, age squared and for the interaction between year of birth dummies and the attendance rate in 1993 in the municipality of residence. Significance levels are denoted as follows: * p<0.10, ** p<0.05, *** p<0.01. Sample: Rural women aged 15-49 years old whose husband is not in the cohort affected by the policy.

Source: DHS Benin 2011 and 2017.

Table A-1.14: Tolerance of Intimate Partner Violence, Non migrant Women

	Goes out without telling husband	Neglects the children	Argues with husband	Refuses sex	Burns the food	IPV index	Second stage IPV index
Panel A: Interest Experiment: Individuals aged 4 to 8 or 12 to 17 in 1997							
High intensity*Treat	-0.107**	-0.121**	-0.097*	-0.106**	-0.026	-0.269**	
	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)	(0.11)	
Primary educ							-2.880
							(1.96)
Controls Individual	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Robust to MHT	Yes††	Yes††	Yes†	Yes††	No	NA	
Mean Dep. Var.	0.29	0.31	0.29	0.19	0.18	0.06	0.06
N	1782	1782	1782	1782	1782	1782	1782
r2	0.30	0.19	0.21	0.14	0.14	0.22	-0.67
F	2.23	1.77	2.43	3.09	1.24	1.68	1.5e+06
Panel B: Placebo Experiment: Individuals aged 13 to 17 or 21 to 26 in 1997							
High intensity*Placebo	0.077	0.038	-0.008	0.080	-0.035	0.089	
	(0.07)	(0.07)	(0.06)	(0.06)	(0.06)	(0.14)	
Primary educ							-3.778
							(9.17)
Controls Individual	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Robust to MHT	No	No	No	No	No	NA	
Mean Dep. Var.	0.29	0.30	0.27	0.18	0.17	0.04	0.04
N	1097	1097	1097	1097	1097	1097	1097
r2	0.28	0.21	0.23	0.17	0.19	0.23	-0.86
F	2.15	3.58	2.17	1.92	5.77	2.30	824.54
Regression Kink Design							
Normalized birth year* post kink* High intensity	-0.00563	-0.01051*	-0.00785	-0.00363	-0.00540	-0.01934	
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	
Primary educ							-1.66556
							(1.34)
Controls Individual	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Robust to MHT	Yes‡	No	No	No	Yes‡	NA	
Mean Dep. Var.	0.29	0.30	0.29	0.19	0.18	0.06	0.06
N	2954	2954	2954	2954	2954	2954	2954
r2	0.27	0.15	0.17	0.10	0.11	0.19	-0.08
F	1.65	2.05	1.85	1.87	1.44	1.74	897.01

Note: The dependent variable is in the following order a dummy taking the value 1 if the woman finds wife beating acceptable if a woman goes out without telling her partner, argues with him, neglects the children, refuses sex and burns the food. We control also for ethnicity, religion, age and age squared. Panel A and panel B present the results of the double difference model. Significance levels are denoted as follows: * p<0.10, ** p<0.05, *** p<0.01. Sample: Rural women aged 15-49 years old, who have never migrated or who arrived in their current location before the age of 8.

Source: DHS Benin 2017.

Table A-1.15: Age and education gap with the husband

	Age difference with husband	Education difference with husband
Panel A: Interest Experiment: Individuals aged 4 to 8 or 12 to 17 in 1997		
Number of school built between 1997 and 2003 * Treat	-0.270	-0.001
	(0.25)	(0.02)
Controls Individual	Yes	Yes
Mean Dep. Var.	7.97	0.29
N	5333	5333
r2	0.05	0.11
F	1.91	2.26
Panel B: Placebo Experiment: Individuals aged 13 to 17 or 21 to 26 in 1997		
Number of school built between 1997 and 2003 * Placebo	0.334	0.045***
	(0.27)	(0.02)
Controls Individual	Yes	Yes
Mean Dep. Var.	8.38	0.29
N	4741	4741
r2	0.07	0.13
F	2.79	3.63

Note: The dependent variable is the the difference between the age of the husband and the age of the bride, for the first column, and the difference in attendance to primary school for the second column. All specifications include municipality dummies, year of birth dummies and interactions between the year of birth dummy and the number of children in the district of birth in 1993. We control also for ethnicity, religion, age and age squared. Significance levels are denoted as follows: * p<0.10, ** p<0.05, *** p<0.01. Sample: Married rural women aged between 15 and 49.

Source: DHS Benin 2011 and 2017.

Table A-1.16: Respondent's say in decision-making

	some say in decision	decides alone	husband decides alone
Panel A: Individuals aged 4 to 8 or 12 to 17 in 1997			
Number of school built between 1997 and 2003 * Treat	-0.046	-0.005	0.049
	(0.05)	(0.04)	(0.05)
Controls Individual	Yes	Yes	Yes
Mean Dep. Var.	-0.05	-0.05	0.05
N	3421	3421	3388
r2	0.16	0.09	0.16
F	2.12	1.82	2.06
Panel B: Placebo Experiment: Individuals aged 13 to 17 or 21 to 26 in 1997			
Number of school built between 1997 and 2003 * Placebo	0.038	-0.023	-0.040
	(0.03)	(0.04)	(0.03)
Controls Individual	Yes	Yes	Yes
Mean Dep. Var.	0.10	0.06	-0.10
N	3210	3210	3195
r2	0.16	0.09	0.16
F	1.41	1.64	1.42

Note: Women residing in rural areas. The dependent variable in column 1 is a standardized score of women having some say in decision-making (either alone or jointly with the husband). The dependent variable in column 2 is a standardized score of women being the sole decision maker, and the dependent variable in column 3 is standardized score of respondent's husband being the sole decision-maker. All specifications include municipality dummies, year of birth dummies and interactions between the year of birth dummy and the number of children in the district of birth in 1993. We control also for ethnicity, religion, age and age squared. Significance levels are denoted as follows: * p<0.10, ** p<0.05, *** p<0.01.

Source: DHS Benin 2010 and 2017. Women residing in rural areas.

Tolerance of IPV as a relevant measure of marital outcome

We argue along with Hanmer and Klugman (2016) that tolerance of IPV serves as a meaningful proxy for capturing one dimension of women's marital outcomes, as it reveals a facet of their own sense of empowerment and well-being. Despite successes in disseminating the notion of women's empowerment, there persists no clear and consensual definition of the concept. For this reason, we sought inspiration in the development literature's reflections on women's empowerment (Kabeer (2005), see Mosedale (2005) for a review of it) and combined several approaches in our own proposed definition of it. We define empowerment as the process by which women become aware of and challenge the gender norms that curtail the realm of possibilities available to them, as compared to men. We also include their ability to choose and act individually and collectively in pursuit of their own strategic interests. Although this definition is close to that of Kabeer (2005), it also explicitly reintroduces the psychological process of awareness necessary to challenge power relations betweens sexes, which is closer to the definition used by Stromquist (1999).

Starting from this definition, to study tolerance of IPV is a matter of studying woman's individual support of a norm that allows a man (here, a partner or husband) to exercise his physical power (i.e. physical violence) to police a woman's behavior. The DHS questions focus on two factors. First, regarding the scenarios described as perceived justifications for wife beating, the questions presuppose that the behaviors transgresses gender norms. In the present case, the literature on Western African societies has largely documented that the behaviors mentioned refer to what is indeed expected from women (though it is less clear for the item "burning the food"). Second, if the behavior mentioned is considered transgressive and a woman finds it acceptable for a husband to beat his wife in those circumstances, this actually means that a woman recognizes the husband's prerogative to police wife's behavior, which, in other words, means that they have legitimate power over them that impedes their right to physical integrity.[47] Additionally, the literature has shown that, in some contexts, women's individual and collective tolerance of IPV was positively associated with the risk of experiencing IPV (Boyle *et al.* (2009) in India). In Jewkes (2002), the author relies on a cross-cultural analysis to demonstrate that the occurence of IPV is stronger in contexts where physical violence against

[47] It could be argued that some women may answer "yes" to this question because they are the senior wives and know that, should violence take place in the household, they would not be the recipients of it. It may also be because the respondent lives with her son and his wife and thus answers the question with her daughter-in-law in mind. In both these examples, the respondents would derive their protection having seniority, not from being a woman. The definition of empowerment that we use, together with that of Mosedale (2005), is not incompatible with women deriving power from their age, ethnicity, wealth or family position. However, this power also is not derived from being a woman, which still says something about their own sense of women's empowerment.

women is condoned.

CHAPTER 2

WITHIN FAMILY DYNAMICS AND INTIMATE PARTNER VIOLENCE (IPV) IN BURKINA FASO

Abstract: I explore whether, in Burkina Faso, having boys strengthens the status of their mother, proxied by women's tolerance of IPV, their autonomy in decision-making, their probability to have anemia, and their experience of IPV. I rely on the exogeneity of the sex of the first child, and of the first two and three children. I find suggestive evidence that the impact of the sex of the first child varies according to the type of union. Among polygamous wives, first wives whose eldest child is a son seem to benefit from more autonomy in decision-making. As for second wives, having a son is correlated with less tolerance of IPV and with an increased access to resources proxied by the probability to be anemic. I also find suggestive evidence that staying too long without a son increases exposure to IPV, but that, for monogamous women, once a son is born, it might also lead to a violent behavior from the husband. The differences in outcomes between polygamous and monogamous wives are difficult to account for. The ambiguity of the results also calls into question the quality of the measure of tolerance of IPV, participation in decision-making, and of experience of IPV used in large household surveys.

Introduction

The interactions between family members in Africa are complex: the prevalence of non-nuclear families and polygamy, in particular in West Africa, stir within-household power dynamics that put an unequal strain on family members according to their gender or their age. The patriarchal organization of many African societies lead to situations where women are dominated by their husband or by his family at different steps of their life cycle. Those unbalanced relationships are notably translated in the fact that most of the victims of domestic violence are women. And the phenomenon is widespread in Africa. In Burkina Faso, according to DHS data for 2010, nearly one third of women believed it was justified for a husband to beat his wife if she goes out without telling him.[1]

Because men remain the pillars of the lineage in patriarchal and virilocal societies, the descendants of a woman are likely to impact her status in the household (Bledsoe *et al.* (1998)). Son preference, though more subtle than in South Asia, has been proven to exist in sub-Saharan Africa; it is notably revealed through patterns of fertility choices of women (Rossi and Rouanet (2015), Lambert and Rossi (2016), and Milazzo (2014)), because having a son provides women with a stronger claim on the husband's resources. Access to resources, however, is only one facet of women's empowerment and status within her household (Miedema *et al.* (2018)), and the impact of having a son on one's mother intrinsic agency, proxied by attitude towards wife-beating and experience of intimate partner violence (IPV), has seldom been studied.

The paper addresses whether having a boy as a first child is part of the statutory attributes that reinforce the position of a woman in her family, leading her to accept less domestic violence, or to experience less violence. Using the 2003 and 2010 DHS for Burkina Faso, I exploit the exogeneity of the gender of the first born, and the absence of preferences for boys in fertility behavior in Burkina Faso (Rossi and Rouanet (2015)). I also take advantage of the prevalence of polygamy in Burkina Faso to use a husband-fixed effect strategy that allows me to control for all unobserved characteristics of the household. When the husband-FE strategy does not apply, I use a linear probability model. Finally, I explore the impact of the timing of birth of the first son by using a duration model to predict the probability of still not having a boy at a time t that I then use as an independent variables in regressions whose outcomes are the tolerance of IPV, decision-making in the household, the probability to have anemia, and the experience

[1] This share reached 41% in Senegal (DHS 2014) and 55% in Mali (DHS 2012-2013).

of IPV.

I find suggestive evidence that for women married to a polygamous partner, having a son may lead to improve the mother's agency and, to a certain extent, may shelter her from experiencing IPV. The results I describe have to be understood with caution, as they are not robust to multiple hypothesis testing, and some estimates are too unprecise to allow for a clear-cut answer. I find that for first wives, having a first born son is positively correlated with having more say in the decision-making process compared to first wives whose first child is a girl, and to second wives, irrespective of the sex of their eldest child. When exploring whether the impact of the sex composition of the other spouse matters for women's own agency, I find that being a first wive who has at least one boy among her first three children and whose junior spouse only has girls among her first three children, is positively correlated with making decision on own's own. I also find suggestive evidence that second wives whose first two children are girls may incur a penalty in terms of tolerance of IPV compared to first wives, and compared to second wives whose first child is a boy. As for the timing of the birth of the first son, it might be that, for second spouses, giving birth to a son at the very beginning of the union is negatively correlated with the probablility to be anemic.

The results for women married to monogamous partner are more ambiguous. I find that the higher the probability to still not have a son at time t, the less they tolerate IPV. I also find that being the monogamous mother of a first child who is a boy is correlated with more emotional and physical violence, and that having two girls or three girls among the first two or three children is associated with less controlling behavior. The study of the timing of the birth of the first son might nuance this finding, as the probability to still not have a son at time t is positively correlated with experiencing controlling behavior among monogamous women. As for women married to a polygamous partner, I also find that having two or three girls among the first two or three children is associated with less IPV, but the results also suggest that the higher the probability to still not have a son at times t, the more likely they are to experience physical IPV. It might suggest that not having a son exposes women to IPV, but that once a son is born it might trigger a more controlling behavior by the husband.

The qualitative differences between polygamous and monogamous wives are puzzling. The difference could come from the absence of a household fixed-effect in the regression for monog-

amous women; if the intensity of violence is so intense that mothers of a first born daughter have to leave their partner, then women can stay with a violent partner only if the firt born is a boy. The complexity of the descriptive results also calls into question the quality of the measure of tolerance of IPV, participation in decision-making, and experience of IPV used in large household surveys.

This paper speaks to the literature that studies the impact of family structure on women's tolerance and experience of IPV, and suggests that the impact of having sons differ according to the type of union women are in (monogamous or polygamous), to their rank among the cospouses, and to the timing of birth of the first boy. For monogamous women having sons seems to be mainly correlated with negative outcomes. The result is compatible with Yount (2005) who analyzes the impact of the family setting on domestic violence in the district of Minya in Egypt, and finds that having sons enhances the vulnerability of women. In the same vein as Heath and Tan (2018) who shows that in Bangladesh and India having daughters increases their mother's participation in household decisions, I find that having daugthers is negatively correlated with IPV among monogamous women, and to a certain extent among polygamous women. Most importantly, I contribute to the literature by showing that the impact of the gender composition of children on their mother's tolerance of IPV, decision-making, and experience of IPV are not aligned. The effect of gender composition appears to be complex and multi-faceted, but may also suggest that the way these phenomena are measured in large household surveys is flawed. Finally, the paper contributes to the larger literature on the determinants of IPV (Alesina *et al.* (2016), Flatø and Kotsadam (2014), Bobonis *et al.* (2013), Anderson and Genicot (2015)).

The paper is structured as follows: Section 2.1 and section 2.2 provide context on family structure in Burkina Faso, and describe the data. Section 3.3 describes the empirical strategy and section 3.6.1 presents the results. Section 2.5.4 discusses the results and Section 2.6 concludes.

2.1 Context

2.1.1 Preference for boys in Africa

In the African context, family ties are key because its members rely on one another to provide them with an informal safety net they couldn't find elsewhere. Moreover, the majority of the ethnic groups found in Africa are patrilinear and virilocal. In this regard, Burkina Faso is no exception. As put forward in Rossi and Rouanet (2015), based on classification by Gray (1998),

the three groups identified as matrilinear in Burkina Faso and coded in the DHS represent less than 5% of the population. Within patrilinear and virilocal systems, women's status remain more precarious than the one of their male counterparts.

"In the traditional family organization in Africa, women are both respected because they are the depositaries of the power of reproduction, and constantly "controlled" because of this "dangerous" priviledge. Parentality systems determine women's status and their access to the resources of both their own families and their spouse's. It is usually through their children that women are allowed to access the means of production of their spouse's family, yet this access remain precarious. [...] They are maintained in a status of inferiority [...]. The least priviledged are the youngest women. They depend on both the men of their own family or their husband and on senior women: their mother-in-law, in particular [...]." (Locoh (1995), p.25).

In this type of social organizations, transmission of land property goes through male heirs. This aspect is enhanced in the Burkinabe context where the majority of people are muslim, because the islamic inheritance law favours boys compared to girls.

There have been debates on the existence of a preference for boys in the African context. Deaton (1989) found no discrimination against girls in household expenditure data. However, a more recent literature shed light on other areas where preference for boys could find an expression. A significant body of the literature provided evidence that the fertility behaviour of women revealed a discrimination against girls. The anthropological literature pinned down the phenomenon.

"In their husbands' compounds, women seek to establish their security and to gain a competitive edge over present and future co-wives and sisters-in-law by bearing a number of children, especially sons, who will retain rights of residence and inheritance in the compound and will eventually take over its leadership roles." (Bledsoe *et al.* (1998), p.23).

The economic literature also showed evidence of both the importance of having a son (Rossi and Rouanet (2015), Milazzo (2014)) and of a rivalry between children of wives, both theoretically (Rossi (2019) between co-wives) and empirically as did Lambert and Rossi (2016) (between children of a man born from different monogamous unions). The authors demonstrate that, in Senegal, women at risk of becoming widows and whose husband already have children from a previous union, increase their fertility until having a boy, which will allow them to secure residence and access to property, should their husband die. In Burkina Faso, however, Rossi

and Rouanet (2015) do not find evidence of a differential birth spacing behavior according to the sex of the children. It suggests, that in this context, the sex composition of children does not trigger a change in fertility behavior of women.

2.1.2 Sex of the first born child

In a context of high fertility, all women end up having at least one boy. As result, I would not have enough variation in the data to assess the effect of having a boy on women's tolerance of IPV or domestic violence. The first born child however, is important in its own right. It corresponds to the first experience of motherhood of a woman. She gives birth to the eldest of the couple, the one that arrives rather rapidly after the beginning of the union. Also, in the absence of sex-selective abortion, which is arguably the case in Burkina Faso,[2] the gender of the first child is exogenous to violence. It is for all these reasons, that when I investigate whether the sex composition of the children has an impact on the acceptance and experience of domestic violence of women, I favor the study of the impact of the gender of the first born child. Burkina Faso, however, is a high fertility context, as women had on average 5.87 children in 2010,[3] which means that the gender of the first two or three children is likely to be exogenous.Rossi and Rouanet (2015) make the case for the exogeneity of the gender of the first two or three children as it does not find evidence of a differential birth spacing behavior according to the sex of the children. Using the DHS, I also perform a raw exploration of whether there exists a birth stopping rule based on the gender of the children. If such a rule existed and favored boys, I would expect to see that girls are born at relatively earlier parities than boys. As shown in table A-2.3 in the appendix, there is no evidence of a birth stopping rule favoring boys. This raw test is in line with the findings of Basu and De Jong (2010), who find that there is no evidence of a birth-stopping rule that would reflect a preference for sons in Burkina Faso. Since there is no evidence of birth-spacing or birth-stopping rules, I also explore the impact of the birth parity of boys on their mother's outcomes.

2.1.3 Declaration of the sex of the first child

The data collected on the sex of the first child is reported by women themselves. It may be an issue if women remember short-lived birth differently according to the gender of the child.

[2]In Burkina Faso, there are no widespread means of knowing the gender of a foetus such as ultra sound echography for instance. A a result I believe it is reasonable to assume that selective foetus abortion is not a threat to the identification strategy.

[3]Source: World Bank databank.

Indeed, it has been documented that in contexts of high fertility, women do not necessarily recall accurately the births of all their children. It may be particularly true for the oldest respondents of my sample. If this recall bias existed and wasn't random, then the sex of the first born wouldn't be as good as randomly assigned. Additionally, if short-lived female births were endogenous to IPV, it would lead to ignore the effect of the differential effect of the sex of the first born on domestic violence. To check whether it is an issue for the analysis, I compute the sex ratio of all the births enumerated in the dataset, for the whole sample. It is 105.16,[4] which is a standard sex ratio. Table A-2.4 in Appendix provides sex ratios of all children born broken down by the age category of the mother. It shows that the sex ratio are rather balanced across age groups, suggesting that there may not be issues of recall bias from older woman, who may remember less dead infant girls than dead infant boys, nor of selective mortality at higher ages, which could have been the sign that mothers of boys have a better chance to survive until older ages than mothers of girls.

2.2 Data and Descriptive Statistics

2.2.1 Data

I use the 2003 and 2010 Demographic and Health Survey (DHS) of Burkina Faso, that survey women aged 15 to 49 years. All women are eligible to answer questions about the tolerance of domestic violence. In the 2010 survey, one eligible woman per household is randomly selected to answer a specific module on the experience of violence. Additionnally, in the DHS 2010, all men aged 15-59 years old were surveyed in half of the selected households, and in the DHS 2003, all men aged 15-59 years old were surveyed in a third of the selected households. In order to account for the changing population structure between 2003 and 2010, the survey weights were denormalized.[5]

The DHS is widely used in the literature on empowerment and domestic violence because it is one of the very few surveys that includes questions on women's attitude towards domestic violence and on their actual experience of domestic violence. The DHS is particularly suited to my study as, along with measures of domestic violence, it also provides the complete birth history of women surveyed. It also has the advantage of allowing to match some women with

[4] author's calculation from DHS 2003 and 2010 Burkina Faso for women in their first union who had at least one child.

[5] I am grateful to Juliette Crespin-Boucaud (PSE) for sharing her codes to denormalize the DHS weights.

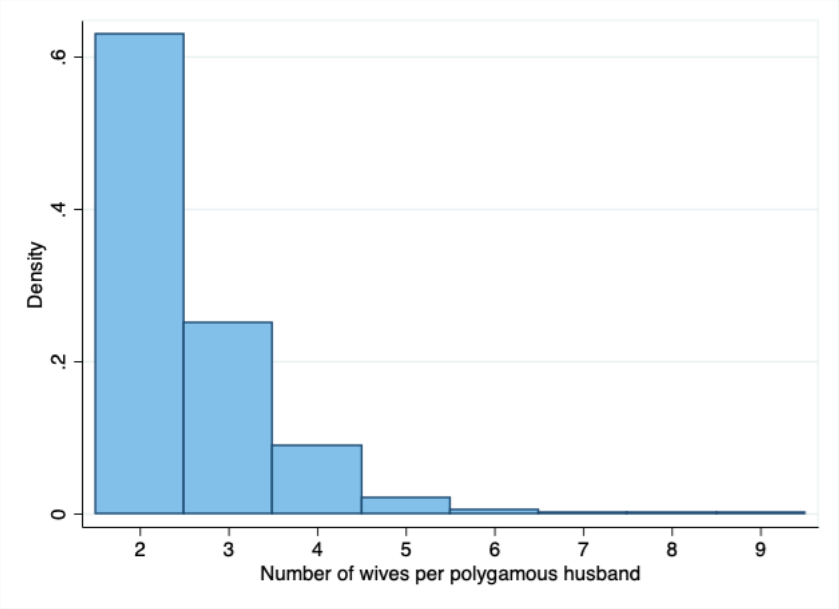
their partners when the former is surveyed. However, the dataset also comes with a few shortcomings for the purpose of this study. First, the eligibility rule of women narrows down data collection to individuals aged 15 to 49 years. As a consequence, in the case of a polygamous union, if one or several wives are older than 49 years old, I am unable to access part of the information concerning this family cell. The second shortcoming lies in the definition of the household chosen for the DHS. A household is defined as people who live under the same roof or who spent the night before the day of the interview. This definition doesn't allow to identify polygamous women married to the same partner but who do not reside in the same house. As a result, it is likely that co-residing polygamous women be over-represented in the sample compared to their presence in the whole population, and it is very uncertain how this feature could bias my results.

Childlessness is not studied in the paper. Given the importance of motherhood in the studied context, entering motherhood may play a role no less important than the gender of the eldest child on women's status. This aspect is put aside, however, in order to focus solely on the impact of the gender composition of children on their mothers. Once my sample is restricted to women in first union who had at least one child, the median age at first marriage is 17. As a result, women below this age and who have at least one child are likely to have particular unobserved characteristics that may bias my results. To avoid that, I drop observations for women below 17.

Table A-2.7 in the Appendix summarizes the main characteristics of women in my sample. On average, 75% of respondents lived in a rural area at the time of the survey. 46% of them are muslims. The majority of them belong to the Mossis, the main ethno-linguistic group in Burkina Faso. Mothers of a first born girl are slightly more represented among the Mossis (by 1pp) but I control for the ethno-linguistic group of respondents in the regressions. The respondents are on average 32 years old. 88% of them are in their first union. They are mainly uneducated (84% of them) and in unions with partners who are uneducated as well (81%). They entered their first union at nearly 18 years old, with partners on average 10 years older than them. The first child to be born from this first union arrived a year and a half after the beginning of the union. They gave birth on average to 4 children.[6] 46% of respondents are married to a polygamous partner. Figure 2.1 shows the distribution of cowives per polygamous union: 63% of polygamous unions involves two wives, and 24% of polygamous unions involve three wives.

[6] In 2015, the number of children per woman in Burkina Faso was 5.44 according to the World Bank.

Figure 2.1: Distribution of polygamous wives by number of cowives



Source: DHS Burkina Faso 2003 and 2010. Sample: women aged 17-49 years old and married to a polygamous partner.

2.2.2 Measures of violence and agency in the DHS

Attitude towards domestic violence and decision-making in the household In the DHS, data are gathered on whether women believe it is justified for a husband to beat his wife trough a series of five questions: "According to you, it is justified for a husband to hit or beat his wife in the following situations:

- if she goes out without telling him?
- if she neglects the children?
- if she argues with him?
- if she refuses to have sex with him?
- if she burns the food?"

I build an index of the tolerance of IPV by using a principal component analysis (PCA), and I rescale the score obtained so that it takes positive values (ranging from 0 to 2.78).

I use women's attitudes towards gender violence as a proxy for women's status in the household. Whether women justify wife beating or not has been validated as a measure of women's intrinsic agency, that is defined in Miedema *et al.* (2018) as "the extent to which women's ex-

pression of gender attitudes reflect or reject normative beliefs", in the context of Egypt (Yount *et al.* (2016)) and East Africa (Miedema *et al.* (2018)) notably. There are however limitations to using tolerance of IPV that this literature does not address and that I discuss in Section 2.5.4.

I also explore another aspect of women's autonomy: the reported participation in decision-making in the household, which is identified by Miedema *et al.* (2018) as a measure of instrumental agency. In the DHS, women are asked who has the final say on spendings for the respondent's healthcare, on making large purchases for the household, on purchases for the household daily needs, on visits to family or relatives, and on the food to be cooked each day. The respondent may answer that she is the only one deciding, that she and her partner are deciding jointly, that her partner is the only one deciding, or that someone else is deciding. For this study, I will use answers the identity of the decision-maker for spendings for the respondent's healthcare, for making large purchases for the household, and for visits to family or relatives because they are the three questions that are available across both DHS survey waves. Following standard practice, I build an index of being the sole decision maker using a principal component analysis (PCA); I choose the first loading factor and I rescale the score obtained so that it takes positive values.

The tolerance of IPV, and to a certain extent the report of who takes decision in the household, are likely to be subject to a declaration bias that brings nonrandom noise to its measure. The bias would materialize if a respondent wants to appear loyal to her husband and declare it is tolerable to beat one's partner despite disapproving it, or conveying that he is the one in charge of the household by reporting that he makes most of the decision alone whereas they decide jointly.[7] Alternatively, and despite the efforts made in the methodology of data collection on violence in the DHS to limit this issue, a social desirability bias that may lead women to tell the DHS surveyor they condemn domestic violence for any motive because they think it is what they want to hear cannot be completely ruled out. This is the reason why I use an objective measure of women's health: whether women have mild or severe anemia. Anemia is an abnormal drop in the count of red blood cells. Its main symptom is chronic fatigue. The causes of anemia vary with age but its appearance can be limited by the diet of individuals. I use anemia as a proxy for women's access to a diet that covers their physical needs that may differ according to their access to resources. The DHS meausures anemia by doing a blood test on a subsample of respondents.

[7] Annan *et al.* (2021) documents the disagreement between spouses on the identity of the main decision-maker in the household using the DHS for 12 Sub-Saharan countries, including Burkina Faso.

Experience of violence The domestic violence module focuses on actual violence experienced by the woman surveyed at the hand of her partner. It takes into account the multidimensional aspect of violence with questions targeting verbal, physical, sexual and emotional violence. As a reminder, only one eligible woman per household is randomly selected to answer this module of the DHS.

In the first part of the questionnaire, surveyed women are presented with situations that women can face in their every day life and are asked whether the situation apply to their relationship with their partner or husband. To elicit a controlling behavior from the partner of the woman surveyed, the following questions are asked:

1. "Is/was your partner jealous or angry if you speak/spoke to other men?
2. does/did he accuse you often to be unfaithful?
3. does/did he allow you to meet with your female friends?
4. does/did he try to limit your contacts with you family of origin?
5. does/did he insist on knowing where you are?
6. does/did he not trust you with money?
7. does/did he prevent you from working?"

A short second set of questions targets emotional abuse. The measure of emotional violence boils down to two questions in the DHS:

1. Did your husband ever do or say something to humiliate you in front of other people?
2. Did your husband ever threatened you or someone close to you?

Based on these two questions, a dummy is defined to assess emotional violence. This variable is equal to 1 if a woman answered "yes" to at least one of the items and 0, if she answered "no" to both of them.

Then, a third part of the module, inspired from the Scale of conflict tactics developed by Straus (1990), asks whether the woman was the victim of different type of physical violence at the hand of her partner. The experience of violence is examined starting with the least severe acts of violence ("Has it ever happened that your husband slap you, shook you or threw something at you"), moving to increasingly intense forms of violence ("did he ever punch you", "did he ever dragged you on the ground or kick you?") to reach the most severe forms of physical violence ("did he ever attack you with a knife, a gun or another type of arm?").

I build a score of controlling behavior and a score of physical violence using a PCA. I choose the first loading factor and rescale the scores obtained so that it takes positive values.
As with the tolerance of IPV and decision-making, issues of declaration bias are likely to arise for the declaration of the experience of IPV. First, despite the efforts made in the design of the survey on domestic violence, IPV is likely to be underestimated, especially for the most severe forms of physical violence, and for marital rape (Deschênes (2021), Cullen (2020)). In particular, Deschênes (2021) shows that the underestimation is not random, and that mothers of a first born daughter are more likely to underdeclare IPV than mothers of a first born boy. Because of the limitations, the results on experience of IPV have to be interpreted with caution.

Second, the bias may be even more acute than for tolerance of IPV because experience of IPV is likely to be endogenous to the tolerance of violence. In case a respondent is beaten by her partner, there are three scenarios for women who survived their partner's violence at least until the survey: 1) the respondent can leave her violent partner, i.e. taking the "exit" option; 2) the respondent stays with her abusive partner because she cannot leave him but condemns IPV; 3) the respondent stays with her abusive partner and is so used to violence that she declares she condones it. The weak but positive association between tolerance of violence and declaration of violence displayed in table 2.2 is consistent with the three scenarios.

2.2.3 Descriptive Statistics on agency and on domestic violence

Table 2.1 presents average tolerance and experience of different types of intimate partner violence. The upper part of the table shows that 43% of women find violence acceptable for going out without telling the husband, neglecting the children, or arguing with the partner. Those three reasons for condoning wife beating are more accepted than the ones justified on the grounds of refusing sex or burning the food. These last two items are seen as acceptable justifications for violence by respectively 29% and 17% of respondents. The mean value of the index of tolerance of IPV is 0.97 and the mean value of the sole decision-making index is 0.68. The probability to be anemic is 52% in the whole sample.

It is important to note that the 2003 survey wave included a preambule to the tolerance of IPV questions that read as follows:"Sometimes a husband is upset or angry because things his wife did". The preambule introduces a disculpation of wife beating by suggesting that the wife may be guilty of something, or by normalizing the anger of the husband. The preambule was

dropped in the next survey wave. It may partially explain the difference in the level of tolerance of IPV between 2003 and 2010, shown in figure 2.2. As a result, in the analysis of the impact of the sex composition of children, I systematically include a survey wave fixed effect.

As shown in the lower part of table 2.1, 60% of respondents experienced a form of control from their partner, and 9% of them suffered from emotional violence. 11% of respondents declare that they experienced at least one form of less severe physical violence. Severe physical violence and marital rape are the least declared form of IPV with 2% and 1% of women revealing that they went throught it.

The matrix of correlations between the items of tolerance and the incidence of IPV displayed in table 2.2, confirms the intuition of table 2.1: the first three items of tolerance of IPV are strongly and positively correlated. The correlation is a slightly less strong between the first three items and refusing sex and even less strong for burning the food. It is also interesting to note that items of the experience of violence are positively though weakly correlated (less than 10%) with items of tolerance of IPV. Although more work should be devoted to measures of violence itself, it seems that measures of women's status are little correlated with the declaration of violence. Having said that, it doesn't rule out a declaration bias in the experience of violence according to the empowerment status of respondents.

Table 2.3 shows whether answers of the first wife and the second wife in polygamous household are correlated. The coefficients on the diagonal vary from 31.91% to 39.04%, suggesting that the answers of cospouses are moderately correlated. Finally, I describe the correlation between husband and wife's tolerance of IPV. Table A-2.1 in Appendix shows that there is little to no correlation between the attitude towards IPV of polygamous husbands and their wives, and table A-2.2 in Appendix also suggest that the attitude towards IPV of husband and wife are significantly yet weakly positively correlated.

Table 2.1: Summary statistics on tolerance and experience of IPV

VARIABLES	mean	sd	min	max	N
Measures of Agency and Tolerance of IPV					
goes out w/o telling	0.43	0.49	0.00	1.00	22,821
neglects the children	0.44	0.50	0.00	1.00	22,821
argues with husband	0.44	0.50	0.00	1.00	22,821
refuses sex	0.29	0.45	0.00	1.00	22,821
burns the food	0.17	0.38	0.00	1.00	22,821
sum item tolerance IPV	1.77	1.81	0.00	5.00	22,821
index tolerance IPV	0.97	1.01	0.00	2.77	22,821
index tolerance husband	0.91	1.58	0	6.27	6,928
index sole decision-making	0.68	1.29	0.00	5.50	22,821
anemia	0.52	0.50	0.00	1	9,677
Experience of IPV					
emotional violence	0.09	0.29	0.00	1.00	9,189
less severe IPV	0.11	0.31	0.00	1.00	9,189
severe IPV	0.02	0.14	0.00	1.00	9,189
marital rape	0.01	0.12	0.00	1.00	9,189
control	0.60	0.49	0.00	1.00	9,189
score control	1.05	1.41	0	9.25	9,189
score physical ipv	0.42	1.60	0.00	18.89	9,189

Note: Sample: Women aged 17-49 years old. Source: DHS Burkina Faso 2003 and 2010.

Figure 2.2: Tolerance of IPV across DHS survey waves

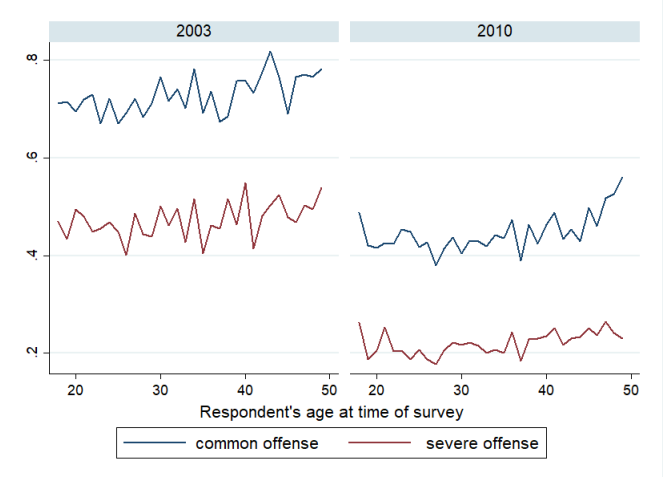


Table 2.2: Correlation of tolerance and experience of IPV

	goes out w/o telling	neglects children	argues with husband	refuses sex	burns the food	index control	emotional ipv
goes out w/o telling	1						
neglects children	0.6555*	1					
argues with husband	0.6086*	0.6490*	1				
refuses sex	0.4636*	0.4768*	0.5141*	1			
burns the food	0.3946*	0.4061*	0.3904*	0.4922*	1		
index controling behavior	0.0419*	0.0558*	0.0483*	0.0260*	0.0392*	1	
emotional ipv	0.0482*	0.0646*	0.0399*	0.007	0.0183	0.2384*	1
index physical ipv	0.0258*	0.0453*	0.0085	0.0349*	0.0115	0.2114*	0.4652*

Note: Sample: Women aged 17-49 years old and randomly surveyed for the domestic violence module. Source: DHS Burkina Faso 2010. * denotes significance level of 5% or less. Adjustment for MHT by Bonferroni method.

Table 2.3: Correlation of tolerance of IPV among cowives

	tolerance of IPV - first spouse				
	goes out w/o telling	neglects children	argues with husband	refuses sex	burns the food
tolerance of IPV - second spouse					
goes out w/o telling	0.3784*	0.3164*	0.2971*	0.2530*	0.2122*
neglects children	0.3451*	0.3822*	0.3328*	0.2669*	0.2222*
argues with husband	0.3301*	0.3286*	0.3815*	0.2549*	0.1951*
refuses sex	0.2705*	0.2646*	0.2837*	0.3904*	0.2570*
burns the food	0.2166*	0.2063*	0.1690*	0.2379*	0.3191*

Note: Sample: Women aged 17-49 years old, married to a polygamous partner with exactly two wives. Source: DHS Burkina Faso 2003 and 2010. * denotes significance level of 5% or less.

2.3 Discussion of the impact of the sex composition of children on women's status, agency and empowerment

The question of interest is whether having a first born son improves the status of the mother within the household. As detailed in section 2.1, when a woman has a son she contributes to the lineage of her husband, and it is expected to strenghten her footing in the household. A change in status could then translate in her feeling confident enough to condemn harmful gender norms like the justification of wife-beating.

In addition to a change in status, having a son may provide women with more access to resources in the household, both at the time of the survey and later in her life cycle. The change could be captured by a greater say in the decisions made in the household, and in the measure of anemia. As for access to future resources, in case of death of the partner, section 2.1 describes that having boys provide women with a stronger claim on the resources of the late partner, and that in polygamous households, it can even lead to a strategic intensification of fertility among cowives (see Rossi (2019) in Senegal). Also, once the first born child is old enough, having an adult son could be expected to improve the mother's bargaining power; having an adult son may improve the threat point of the mother, as she can credibly claim that she would leave her

current partner to reside with an adult son.

As for mothers of first born daughters, the literature has shown, in western countries, that the sex composition of children affects their parents' opinions (Downey *et al.* (1994)) and that having daughters enhances the preoccupations of parents about gender equality (Washington (2008)). Heath and Tan (2018) shows that in Bangladesh having daughters increases women's participation in decisions made in the household. If a similar behavior arises in West Africa then mothers of first born daughters may become more sympathetic to gender equality and declare that they tolerate less wife-beating, or try and be more involved in decision-making to make sure they secure enough resources for their daughter.

It remains an open-question whether in polygamous households, the competition between co-wives for access to resources in the household also translates in terms of competition for status. Most of sub-Saharan societies follow a hierarchical order based on gender and age. Being the first (and older) spouse probably grants status in itself to women. Junior wives may be able to improve their status relative to the senior wife by having a first born son, but the impact of the change in status may also differ according to the gender composition of the children of the senior wife. Similarly for senior wives, the gender of the first child of the junior wife might modify their status in the household.

2.4 Empirical Strategy

My empirical strategy relies on using the gender of the first born child as a shock or disturbance in women' status that is exogenous to domestic violence.

I perform a first check of whether the gender of the first child is as good as random. The results of this balancing test is displayed in table A-2.7 in Appendix. It shows that, overall, the characteristics of mothers of a first born boy and a first born girl are balanced. The age at first cohabitation is slightly higher for mothers of a first born girl than mothers of a first born boy. It is also worth noting that, mothers whose first born child is a boy seem to be slightly represented in the Mossi ethno-linguistic group than mothers whose first child is a girl (50% vs. 52%). I control for those imbalances in my estimations.

In this analysis, I choose to use the gender of the first child, irrespective of his survival status even though under-five mortality in Burkina Faso is 89 per thousand.[8] It is probable that the gender of the first born has an impact only if the child survives especially if the effect of the gender is channeled through insurance against widowhood. On the other hand, it is also possible that the birth of a boy, however short his survival is, is enough to influence the belief of a husband in the capacity of his wife to give birth to other boys in the future and change her status in the household. Even though it seems relevant that only surviving children influence the status of the mother, the survival of the children is not exogenous and are correlated to parents' choice in terms of length of breastfeeding or nutrition for instance, which could reflect a gender preference. As a result, I choose to use the sex of the first child regardless of his survival status.

To examine the impact of the gender of the eldest child on tolerance of IPV, I estimate two specifications. First, the preferred strategy consists in comparing women's response to the gender of their first child on their tolerance of domestic violence, for women married to the same partner. I run husband fixed-effect regressions of several measures of acceptance of violence on whether a woman's eldest child is a boy. Denote the outcome for individual i married to husband z as y_{iz} and having a first born boy as boy_i. The husband fixed-effect regression is:

$$y_{iz} = \alpha + \beta * boy_{iz} + \gamma * X_i + \delta_z + \epsilon_{iz}, \tag{2.1}$$

where δ_z is a husband fixed-effect, X_i a set of individual controls for the respondent, which includes the age category of the respondent, her age at first cohabitation, whether she is a muslim, whether she belongs to the Mossi ethno-linguistic group, the duration of the cohabitation with her partner, her education level, her age difference with her partner, her spousal rank and the total number of births. Standard errors are adjusted for clustering at the husband level. When estimating equation (2.1), the identification assumptions are that i) the sex of the first child of a respondent is exogenous, ii) beyond the additional controls added, there is no difference in the unobservable characteristics of women married to the same man. The total number of births, however, which may proxy for women's intended fertility, is something the FE does not control for. This is why I control for the total births of women, so that I can block the potential impact of the sex of the first child on the mother's subsequent fertility as a channel linking the sex of the first child to the tolerance of IPV. Even though total fertility could be seen as a bad

[8] UNICEF - Estimates generated by the UN Inter-agency Group for Child Mortality Estimation (IGME) in 2015

control as it may be impacted by the sex of the first child, Rossi and Rouanet (2015) shows that the sex of composition of children does not impact birth spacing in Burkina Faso. In addition, and as mentionned earlier, I show in table A-2.3 in the appendix that there is no evidence of a birth stopping rule favoring boys, which is in line with the findings of Basu and De Jong (2010).

When focusing on women married to a polygamous partner, I study whether the impact of the sex of the first child varies according to the rank of the mother. I estimate the following model:

$$y_{iz} = \alpha + \beta * boy_{iz} + \theta * 2ndspouse_{iz} + \rho * boy_{iz} * 2ndspouse_{iz} + \gamma * X_i + \delta_z + \epsilon_{iz}, \quad (2.2)$$

where δ_z is a dummy variable equal to 1 if the respondent is the second spouse of the male respondent z, 0 if she is the first spouse.

Even though the husband fixed-effect is the preferred strategy, it is demanding on the data. First, It can only be applied to women who are married to a polygamous partner. Analyzing the impact of the first born on women in a monogamous women must be done with another strategy. Second, as underlined earlier in section 2.2, applying the husband fixed-effect model requires that I have information on at least two co-wives whose partner was interviewed in the man survey, so that I have an identification number for him in the dataset. Because of this feature, I can only apply the husband fixed-effect on a subsample of women in a polygamous union, which results in a loss in statistical power. Third, for safety reasons, only one woman per household is randomly selected to answer the domestic violence module, which makes it impossible to apply the husband fixed-effect model to outcomes relative to the experience of violence. It is for all these reasons that I also use a linear probability model (LPM) to estimate the impact of the gender of the first child on the tolerance and the experience of domestic violence. Using the same notation that in the husband-fixed effect model, I estimate:

$$y_i = \alpha + \beta * boy_i + \gamma * X_i + \epsilon_i, \quad (2.3)$$

where X_i a set of individual controls for the respondent, which includes the age category of the respondent, her age at first cohabitation, her religion, the duration of the cohabitation with her partner, her education level, her age difference with her partner, her ethnic group, the education level of her partner, their region of residence, the wealth quintile of the household, whether

she is in a polygamous union when applicable, and the rank of the spouse if she is married to a polygamous partner. When focusing on the subsample of women in a polygamous union, I additionally control for the rank of the spouse. Standard errors are adjusting for clustering at the husband level to account for the level of polygamy, or at the DHS cluster level when a husband could not be identified or for women in a monogamous union.

Finally, I will also analyze the impact of the sex composition on first and second wives separately. I use the following specification:

$$y = \alpha + \beta * boyfirstwife + \mu * boysecondwife + \theta * boyfirst * boysecond + \gamma * X + \epsilon, \quad (2.4)$$

where $boyfirstwife$ is dummy equal to 1 if the child of the first wife is a boy, 0 if it is a girl, and where $boysecondwife$ is a dummy equal to 1 if the child of the second wife is a boy, 0 if it is a girl.

Finally, in order to exploit all the information on women's birth history, I emulate Boltz and Chort (2019) and use a duration model to predict the survival function of the event "the birth of the first son", then, in a second step, I use the predicted survival function as an explanatory variable. All women who are in a union[9] at time t are at risk of becoming the mother of a boy for the first time. I am interested here in modelling the survival function usually denoted $S(t)$, which is the probability of still not having a son at time t. In duration models, the failure function is defined as follows: $F(t) = P(T \leq t) = 1 - S(t)$, it represents the probability to know the event up to time t. The key concept of the duration analysis is the hazard function, which is defined as $h(t) = lim_{dt \to 0} \frac{P(T \leq t | Tt+dt)}{dt}$.

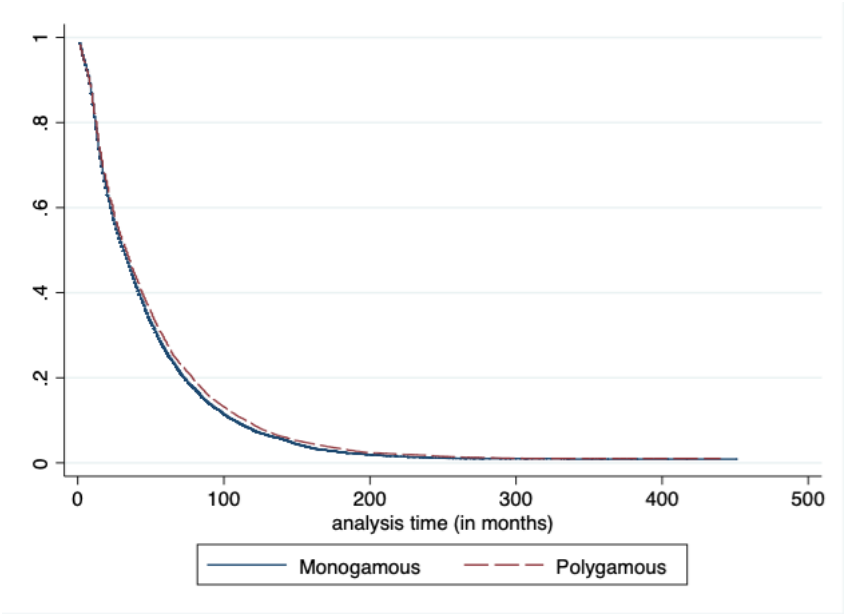
$\frac{P(T \leq t | Tt+dt)}{dt}$ is the probability that the event happens between time $[t; t+1]$ conditional on not having happened at time t. $h(t)$ can be rewritten as $h(t) = -\frac{S'(t)}{S(t)}$. The cumulative risk denoted is defined as follows: $H(t) = \int_0^t h(u)\, du = -log(S(t))$. The survival function can then be rewritten as: $S(t) = e^{-H(t)}$.

In this application, I estimate the survival function using a standard parametric Cox model,so hazard rate is defined by $h(t) = h_0(t) * e^{Xb}, (z = 0, 1)$, with z the stratification variable, which is

[9]The DHS only provide the date of the first union, which is the proxy used here. For monogamous women, I show results for all wives, and for wives in their first union. For polygamous wives, as 20% of junior wives are not in their first union, the proxy of the date of the first union is less likely to be acurate. The solutions provided is to run the tests with and without women in their first union and to see whether the results remain the same. It is acknowledged to be a flawed solution because junior wives in their first union are likely to be selected in a non-random way. For the sake of brevity, the results of these tests are mentionned but the tables are not included. They are available upon request.

the union type: monogamous or polygamous, h_0 the baseline hazard that is non-parametrically estimated on the subsamples of monogamous and polygamous wives, and X a set of time-invariant characteristics of the respondent that are determined at the time of the union, namely the region of residence of the respondent, her religious affiliation, her ethno-linguistic group, her education level, and the wealth of the household. Figure 2.3 shows that there seems to be little difference in the probability to still not have a son between monogamous and polygamous wives.

Figure 2.3: Cox stratified survival function by union type



Source: DHS Burkina Faso 2003 and 2010. Sample: women aged 17-49 years old who have ever been in a union.

I predict for each respondent the probability to not have had welcome one's first son at time t, and in a second step, I use the predicted survival function in the empirical strategy presented above. Using te two-step approach with the Cox model allows to account for the fact that not all women had a first child nor a first son at the time of the survey. In the Appendix, I also the study whether the timing of the birth of the first boy matters by analyzing the impact of the rank of birth of the first son on the outcomes of interest. It is not the preferred way to explore the impact of the timing of birth of the first boy though, because it is based on the subsample of women who have at least one boy, and to make the comparisons more meaningful, on the subsample of women who had four children or less, four being the mean number of children in the sample. The choice to restrict to samples according to the number of births may bias the analysis in ways that are difficult to predict, hence the preferred choice to work with the predicted survival function estimated by a Cox Model.

2.5 Results

2.5.1 Preliminary step: union dissolution and sex of the first child

Women who are not in first union make up for 11.8% of women in the sample, and 15.6% of polygamous women. I explore whether women's whose first child is a boy are more likely to be in a first union than women whose first child is a daughter.[10] In table 2.4, I report the impact of having a boy as a first child on women's probability to be in their first union at the time of the survey. Column (2) and (3) of table 2.4 suggests that, overall, there is no effect of the sex of the first child on their mother's probability to be in first union among monogamous and polygamous women. As expected in this context, second and third spouses are less likely to be in their first union. Column (4) suggests that for second and third spouses being the mother of a first born daughter decreases the probability to be in first union compared to first wives. Even though the interaction terms in column (4) are not statistically significant, the output of a ttest at the bottom of table 2.4 suggests that the overall impact of having a first son rather than a first born daughter for third spouses is different. The insight is confirmed by table A-2.5 that compares third wives in first union to third wives who are not in first union. The comparison confirms that third wives who are not in first union are significantly less likely to be the mother of first born son. Being a third wife who is not in first union is also positively correlated with justifying wife beating and negatively correlated with a score of decision-making in the household. It is thus unclear how the link between the sex of the first child and union dissolution among third wives may impact this study. I consequently choose to exclude third wives from the main analysis of the impact of the sex of the first born on women's agency. Table A-2.10 and table A-2.11 in the appendix show that there is no significant difference in the probability to be in first union among second wives nor among third wives according the sex of the first two or three children.

[10] Among women who are not in their first union, in 97% of the cases, the current partner is not the father of the first child.

Table 2.4: Being in first union and sex of the first child

dependent variable *Respondent is in her first union*	**All** (1)	**monogamous** (2)	**polygamous** (3)	**polygamous** (4)
Boy first born	0.004	0.001	0.007	0.017
	(0.01)	(0.01)	(0.01)	(0.01)
2nd spouse			-0.200***	-0.186***
			(0.01)	(0.01)
3rd spouse			-0.192***	-0.202***
			(0.01)	(0.02)
Boy first x 2nd spouse				-0.027
				(0.02)
Boy first x 3rd spouse				0.020
				(0.02)
Observations	19150	10371	8776	8776
Mean	0.882	0.913	0.845	0.845
pval test: boy 2nd wife = girl 2nd wife				0.446
pval test: boy 3rd wife = girl 3rd wife				0.076

Note: The dependent variable is the probability to be in first union at the time of the survey. Controls: age, education level of the respondent, education level of the partner, difference in education between partners, age difference between partners, age at first cohabitation, total number of births, DHS wealth quintile, region of residence, being a muslim, being a mossi, and the DHS survey wave. I compute robust standard errors. Significance level: * p<0.10, ** p<0.05, *** p<0.01.s Sample: Women aged 17-49 years old. Source: DHS Burkina Faso 2003 and 2010.

2.5.2 Tolerance of IPV and decision-making

The following results are mainly descriptive as they are not robust to controlling for the false discovery rate (FDR).[11] They nevertheless provide avenues to discuss the challenges arising when studying the relationship between women's agency and the gender composition of their children. The results presented below show the pvalue unadjusted for FDR.

Tables A-2.12 and A-2.13 in the appendix present the estimation of equation 2.1 for women in a polygamous union and women in a monogamous union; they suggest that the current set of data does not allow to reject the null hypothesis that the sex of the first child of respondents has no impact on the tolerance of IPV, the score of sole decision-making or the probability to be anemic of their mother.

In table 2.6, I zoom in on women married to a polygamous partner and explore whether the impact of the sex of the first child differs according to the rank of the spouse. In columns 1, 4,

[11] Michael Anderson's sharpened q-values code was used to test whether the results are robust to controlling for multiple hypothesis testing. The procedure is described in Anderson (2008).

and 7, I report the results yielded by the husband fixed effect (FE) estimation, and in columns 2, 5, and 8, I report the results yielded by the linear probability model (LPM) on the subsample of respondents who contribute to identifying the husband FE model.[12] In columns 3, 6, and 9, I report the estimates obtained by using a LPM on the subsample of wives who do not participate to the identification of the husband fixed-effect. The husband FE estimation is the preferred specification because it controls in a more stringent way for husband and household characteristics than the LPM. The comparison between the estimates yielded by the FE and the LPM on women contributing to the identification of the FE is nevertheless of interest, as it allows to explore whether the LPM may yield a bias estimation of the impact of the sex composition of children on women's tolerance of IPV.

Table 2.5: Descriptive statistics agency

Panel A: all women in a monogamous union				
	Mean	**Std Dev**	**Min**	**Max**
score tolerance of IPV	0.88	0.99	0.00	2.78
score sole decision making	0.63	1.20	0.00	5.50
anemia	0.50	0.50	0.00	1.00
Mean by sex of the first child				
	Girl	**Boy**		
score tolerance of IPV	0.88	0.87		
score sole decision making	0.62	0.62		
anemia	0.5	0.49		
Panel B: all women in a polygamous union				
	Mean	**Std Dev**	**Min**	**Max**
score tolerance of IPV	1.09	1.02	0.00	2.78
score sole decision making	0.60	1.17	0.00	5.50
anemia	0.54	0.50	0.00	1.00
Mean by sex of the first child and rank of the spouse				
	1st wife Girl	**1st wife Boy**	**2nd wife G**	**2nd wife B**
score tolerance of IPV	1.04	1.04	1.05	1.06
score sole decision making	0.61	0.66	0.57	0.62
anemia	0.53	0.56	0.53	0.52

Note: Sample: Women aged 17-49 years old, and married to a polygamous partner who has exactly two wives. Source: DHS Burkina Faso 2003 and 2010.

The estimates displayed in the first three columns of table 2.6 are also imprecisely estimated, and do not allow to provide a clear-cut answer to whether the sex of the first child affects the mother's tolerance of IPV, regardless of the specification used. In particular, the FE estimates would be compatible with an effect ranging from -0.095 to 10.1.[13] Column (4) of table 2.6 shows,

[12] It means that I use the subsample of women whose first child is of a different gender than the one of the co-spouse.

[13] Confidence interval with an α equal to 5%.

however, that having a first born son is positively correlated with the score of sole decision-making of first wives, increasing it by 14% compared to first wives whose first child is a girl. Even though having a first born son rather than a daughter decreases the score of decision-making for second wives compared to first wives, the overall effect of being a second spouse and the mother of a first born boy is null, and the data does not allow to reject the hypothesis that, for second wives, having a first born son rather than a daughter is equivalent (pvalue = 0.42), and hence not different than zero. Finally, I find that, among mothers of a first born son, the part taken in decision-making differs for first wives (as it increases) compared to second wives (for whom the data suggests it stays the same) and that the difference between first and second wives is marginally significant (pvalue = 0.15). I cannot provide any conclusion of the impact of the sex of the first child on the subsample of women tested for anemia, as the estimates remain imprecisely estimated. In the appendix, I use the panel of women married to a polygamous partner with 1 or 2 co-wives, of rank 1, 2 or 3 (see table A-2.6), which yields similar results. This first set of results is compatible with the hypothesis that having a first born boy may increase the participation in decision-making of first wives. The effect does not seem to be paralleled among second wives.

Table 2.6: Tolerance of IPV for common offense and sex of the first born child among women in a polygamous union

	score tolerance IPV			**score decision making**			**anemia**		
	FE	LPM contrib. to id	LPM not contributing	FE	LPM contrib. to id	LPM not contributing	FE	LPM contrib. to id	LPM not contributing
Panel: Two cowives of rank 1 and 2									
Boy first born	0.003	0.027	0.019	0.088*	-0.008	0.074	0.000	0.034	0.017
	(0.05)	(0.07)	(0.05)	(0.05)	(0.07)	(0.07)	(0.05)	(0.05)	(0.04)
2nd wife	-0.060	0.071	0.035	0.044	0.019	-0.032	-0.076	-0.042	-0.033
	(0.06)	(0.08)	(0.05)	(0.06)	(0.07)	(0.06)	(0.05)	(0.06)	(0.04)
Boy first x 2nd wife	-0.004	-0.040	-0.036	-0.131*	0.052	0.004	0.018	-0.046	-0.033
	(0.07)	(0.11)	(0.06)	(0.08)	(0.12)	(0.08)	(0.06)	(0.08)	(0.05)
Observations	6476	2047	4429	6476	2047	4429	2879	938	1941
Contributing to id	2047			2047			938		
R-squared	0.009	0.126	0.150	0.024	0.038	0.052	0.028	0.043	0.027
Mean control	1.038	1.036	1.039	0.606	0.528	0.654	0.528	0.528	0.528
Husbands	4454	1049	3405	4454	1049	3405	1955	482	1473
Contributing to id	1049			1049			482		
pval test: boy first + boy first x 2nd wife = 0	0.980			0.421					
pval test: boy first + 2nd wife + boy first x 2nd wife = 0	0.277			0.997					
pval test: 2nd wife + boy first x 2nd wife = 0	0.248			0.149					

Note: In columns 1, 4, and 7, I report the results yielded by the husband fixed effect (FE) estimation, and in columns 2, 5, and 8, I report the results yielded by the linear probability model (LPM) on the subsample of respondents who contribute to identifying the husband FE model. In columns 3, 6, and 9, I report the estimates obtained by using a LPM on the subsample of wives who do not participate to the identification of the husband fixed-effect. In the husband FE, I control for women's ethno-linguistic group, religion, age category, the age difference with the partner, the age at first union, the duration of the union, the total number of birth, and the DHS survey wave. For anemia, I also control whether women are pregnant at the time of the survey. I cluster standard errors at the husband level. In the LPM, I also control for the husband's education, the region of residence, and the DHS wealth score of the household. I compute robust standard errors. Significance level: * $p<0.10$, ** $p<0.05$, *** $p<0.01$. Sample: Women aged 17-49 years old, and married to a polygamous partner who has exactly two wives. Source: DHS Burkina Faso 2003 and 2010.

As Burkina Faso is a high fertility context, I repeat the analysis using the sex of the first two children in table 2.7 and the sex of the first three children in table A-2.14. The estimates yielded by the FE estimation are too imprecise to say whether the sex of the first two children impacts the mother's agency. The same issue arises when the LPM model is applied to the subsample of women who do not contribute to the identification of the FE model (i.e. women whose first two children are of the same gender than the ones of her co-wife). I find, however, that being a second spouse and the mother of two first born daughters is positively correlated with the score of tolerance of IPV (+0.13%) compared to a first wife with at least one boy among her first two children. Among second wives, having two eldest daughters compared to having at least one boy is positively correlated with tolerance of IPV (pvalue = 0.098). The data is hence compatible with the hypothesis that second wives whose first two children are girls are worse off than second wives who have at least one boy among their first two children (or than first wives regardless of the gender composition of their children).

Column (8) of table 2.7 shows that, for second wives, the probability to be anemic is negatively correlated with having at least one boy among the first two children compared to being a first spouse with at least one boy (-26%). Even though anemia is positively correlated with having two daughters among the first two children compared to first spouses with at least one boy (+16.8pp), the overall impact of having two first born daughters is not significantly different than zero. Being a second spouse with at least one boy among her first two children, however, is significantly different than being a second spouse with two first born girls (pvalue = 0.085). It suggests that when it comes to anemia, second wives who have at least one son are better off than second spouses with two first born daughters, whose probability to be anemic is equivalent to that of first wives, regardless of the gender composition of their children. As I do not capture any effect with the FE or with the LPM on women contributing to identifying the FE, the results yielded by the LPM on the non-contributing are probably driven by the comparison of households with different unobservable characteristics. Column (9) of table 2.7 also shows that, among women whose cowife has the same gender composition as theirs, being a second spouse with at least one boy among the first two children decreases the probability to be anemic. Overall, table A-2.14 provide similar insights on the impact of the sex of the first three children on women's agency.

Table 2.7: Women's agency and sex of the first two children

	score tolerance IPV			score decision making			anemia		
	FE	LPM contrib. to id	LPM not contributing	FE	LPM contrib. to id	LPM not contributing	FE	LPM contrib. to id	LPM not contributing
Panel: Two cowives of rank 1 and 2									
GG	-0.028	-0.078	-0.045	0.015	0.038	0.096	0.014	-0.053	0.002
	(0.06)	(0.08)	(0.06)	(0.05)	(0.08)	(0.08)	(0.06)	(0.06)	(0.05)
2nd wife	-0.067	0.014	0.005	-0.022	0.113	0.000	-0.062	-0.143*	-0.056*
	(0.05)	(0.10)	(0.04)	(0.05)	(0.11)	(0.05)	(0.04)	(0.08)	(0.03)
GG x 2nd wife	0.021	0.098	0.134*	-0.047	-0.112	-0.117	0.014	0.168*	0.090
	(0.09)	(0.13)	(0.08)	(0.09)	(0.14)	(0.10)	(0.08)	(0.10)	(0.06)
Observations	6242	1365	4877	6242	1365	4877	2772	612	2160
Contributing to id	1365			1365			612		
R-squared	0.011	0.150	0.141	0.020	0.042	0.046	0.033	0.076	0.028
Mean Control	1.045	1.071	1.042	0.633	0.551	0.646	0.550	0.541	0.552
Husbands	4379	704	3675	4379	704	3675	1922	316	1606
Contributing to id	704			704			316		
pval test: GG + GG x 2nd wife = 0	0.928		0.098	0.646		0.764		0.085	0.048
pval test: GG + 2nd wife + GG x 2nd wife = 0			0.105			0.786		0.693	0.479

hline

Note: In columns 1, 4, and 7, I report the results yielded by the husband fixed effect (FE) estimation, and in columns 2, 5, and 8, I report the results yielded by the linear probability model (LPM) on the subsample of respondents who contribute to identifying the husband FE model. In columns 3, 6, and 9, I report the estimates obtained by using a LPM on the subsample of wives who do not participate to the identification of the husband fixed-effect. In the husband FE, I control for women's ethno-linguistic group, religion, age category, the age difference with the partner, the age at first union, the duration of the union, the total number of birth, and the DHS survey wave. For anemia, I also control whether women are pregnant at the time of the survey. I cluster standard errors at the husband level. In the LPM, I also control for the husband's education, the region of residence, and the DHS wealth score of the household. I compute robust standard errors. Sample: Women aged 17-49 years old, married to a polygamous partner and who declare that they are the first wife or the second wife of their partner. Source: DHS Burkina Faso 2003 and 2010.

Results in table 2.6 suggest that the impact of the sex of the first child may differ according to the rank of the spouse. I further explore the question by examining the impact of the sex composition of the children of each spouse on her own tolerance of IPV, and on the tolerance of IPV of the cospouse with a separate linear probability model for each subgroup of wives. As suggested by table A-2.15 and table A-2.16, I cannot exclude that the sex of the first child nor that the sex of the first two children of a respondent (whether her own or that of her cowife) has no impact on women's agency, whether she is the first or the second spouse. As shown in column (3) of table 2.8, I find evidence, however, that for a first wife, having a junior cospouse whose first three children are girls is positively correlated with the index of the participation in decision-making in the household (+32% compared to the mean). Column (5) of table 2.8 underlines that for first spouses, being the mother of three first born girls decreases the probability to be anemic by 17% compared to having at least one boy. The result is not robust to interacting the gender of the first three children of the first spouse with the gender of the first three children of the second spouse. As shown in panel B of table 2.8, I am unable to rule out that there is no impact among second spouses.

In addition to the sex of the eldest children, I explore whether the timing of the arrival of the first boy matters for the status of women in the household; I use a Cox model to predict the probability to still not have a son at time t. This model allows me to keep the whole sample for the analysis and to account for the right censored nature of the data, as some women may not have had a first son yet. I then use the predicted probability of still not having a son at time t as an explanatory variable in a FE regression for polygamous women, and in a LPM for monogamous women. The estimates shown in panel A and column (1) of table 2.9 have large standard errors, and the data is compatible with the increase in the probability to still not have a son at time t to be correlated with a change in the index of tolerance of IPV ranging from -0.07 to 0.20 (confidence interval with an α = 5%). Column (2) of table 2.9 suggests that for first wives, as the probability to still not have a son increases, the change in tolerance of IPV may vary from -12 to 0.18, and that for second wives who virtually welcome their first son right at the beginning of their first union, the index tolerance of IPV may vary from -12 to 0.16. The interaction term suggests that as the probability to still not have a son at time t increases, the difference in tolerance of IPV between second wives and first wives decreases. The insights do not seem to be driven by second spouses whose first child is not the one of their current partner, as the results remain the same when focusing on women in first union only.[14] Columns (6) of table 2.9 suggests that for second wives who virtually have a son right after after the

[14]Tables available upon request.

Table 2.8: Women's agency and sex of the first three children of the cospouse

	index tolerance		index decision-making		anemia	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: First wives						
1st wife's first three children are girls	0.009	0.016	-0.041	-0.041	-0.091*	-0.077
	(0.075)	(0.078)	(0.080)	(0.078)	(0.054)	(0.057)
2nd wife's first three children are girls	0.007	0.016	0.178*	0.178*	0.064	0.084
	(0.088)	(0.093)	(0.105)	(0.103)	(0.070)	(0.076)
1st wife GGG x 2nd wife GGG		-0.089		-0.004		-0.178
		(0.243)		(0.511)		(0.203)
Observations	1755	1755	1755	1755	802	802
R-squared	0.092	0.092	0.042	0.042	0.070	0.071
Mean/Mean control	1.028	1.032	0.557	0.539	0.531	0.535
pval test: 2nd wife GGG = 2nd wife GGG x 1st wife GGG		0.756		0.929		0.182

	index tolerance		index decision-making		anemia	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel B: Second wives						
1st wife's first three children are girls	0.036	0.024	-0.073	-0.066	-0.059	-0.056
	(0.074)	(0.078)	(0.064)	(0.067)	(0.059)	(0.063)
2nd wife's first three children are girls	0.001	-0.016	0.072	0.083	0.093	0.097
	(0.091)	(0.098)	(0.095)	(0.102)	(0.072)	(0.076)
1st wife GGG x 2nd wife GGG		0.169		-0.108		-0.041
		(0.251)		(0.237)		(0.230)
Observations	1761	1761	1761	1761	810	810
R-squared	0.108	0.108	0.085	0.085	0.058	0.058
Mean/Mean control	1.013	1.009	0.514	0.515	0.520	0.524

Note: I control for women's ethno-linguistic group, religion, age category, the age difference with the partner, the age at first union, the duration of the union, the total number of birth, the DHS survey wave, the husband's education, the region of residence, and the DHS wealth score of the household. GGG stands for the first three children of a respondent are girls. Significance level: * p<0.10, ** p<0.05, *** p<0.01. Sample: Women aged 17-49 years old, married to a polygamous partner and who declare that they are the first wife or the second wife of their partner. Source: DHS Burkina Faso 2003 and 2010.

beginning of their union, the probability to be anemic decreases. Panel B table 2.9 suggests that among monogamous wives, the higher the probability to still not have a son at time t, the less the mother tolerates IPV, whether I use the pooled sample or monogamous women who are in their first union only. The tests may be underpowered to detect an impact of the probability to still not have a son on decision-making or anemia.

Table 2.9: Agency and probability not to have a son at time t

	index tolerance IPV		**index decision-making**		**anemia**	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Polygamous wives						
1st wife: time btw first union and first boy	-0.065	0.031	0.036	0.060	0.013	-0.066
	(0.07)	(0.09)	(0.07)	(0.09)	(0.07)	(0.09)
2nd wife	-0.096**	0.024	-0.022	0.008	-0.061	-0.161*
	(0.05)	(0.09)	(0.05)	(0.09)	(0.04)	(0.08)
2nd wife: time btw first union and first boy		-0.210*		-0.052		0.176
		(0.13)		(0.14)		(0.12)
Observations	6405	6405	6405	6405	2854	2854
R-squared	0.012	0.013	0.018	0.019	0.020	0.023
Unconditional Mean	1.035	1.035	0.604	0.604	0.531	0.531

	index tolerance IPV		**index decision-making**		**anemia**	
	all	first union	all	first union	all	first union
Panel B: Monogamous wives						
time btw first union and first boy	-0.088**	-0.074*	0.057	0.067	-0.013	-0.048
	(0.04)	(0.04)	(0.06)	(0.06)	(0.03)	(0.03)
Observations	10563	9652	10563	9652	4560	4137
R-squared	0.129	0.125	0.034	0.034	0.024	0.028
Unconditional Mean	0.869	0.860	0.623	0.614	0.496	0.497

Note: The probability to still not have a son at time t is the predicted survival function to knowing the event "the birth of the first son". It is obtained using a standard semiparametric Cox model stratified by the polygamous status of the respondent. Variables used to predict the survival function are a set of characteristics of the respondent that are determined at the time of the first union: region of residence, wealth of the household, religious denomination of the respondent, and ethno-linguistic group. Results presented in panel A are yielded by the husabnd-FE estimation, whereas results in Panel B are yielded by a LPM. In the husband FE, I control for women's ethno-linguistic group, religion, age category, the age difference with the partner, the age at first union, the duration of the union, the total number of birth, and the DHS survey wave. For anemia, I also control whether women are pregnant at the time of the survey. I cluster standard errors at the husband level. In the LPM, I also control for the husband's education, the region of residence, and the DHS wealth score of the household. Sample of Panel A: Women aged 17-49 years old, married to a polygamous partner and who declare that they are the first wife or the second wife of their partner. Sample of Panel B: Women aged 17-49 years old, married to a monogamous partner. Source: DHS Burkina Faso 2003 and 2010.

The impact of the sex of the first child on first and second wives may be complex to identify because it may modify the status of the wives and, to a certain extent, of the husband at the same time. Then, what may matter for women's own tolerance of IPV may be the extent to which her husband is treated. Table A-2.20 and table A-2.21 in the appendix does not point towards this direction; I find no impact of the sex of the eldest child of the husband on women's tolerance of IPV, participation in decision-making, or anemia, regardless of their rank, or type of union, though the estimates yielded are imprecise.

Finally, I examine whether the husband's own tolerance of IPV is modified by the sex of the children of his wives. As shown in column (1) of table 2.10, the estimates are imprecisely estimated, which does not allow me to conclude whether the tolerance of IPV of husbands is affected by the sex of his first child. Being a polygamous man married to a second spouse who has a first born son and to a first wife who does not have a first born son is negatively correlated with tolerance of IPV (-41% compared to hubands whose two wives' eldest children are girls).

Table 2.10: Tolerance of IPV of husbands

Dependent variable: husband's tolerance of IPV	(1)	(2)
Panel A: Polygamous husbands		
Husband's eldest is a boy	0.047	
	(0.123)	
1st wife's child is a boy		0.089
		(0.206)
2nd wife's child is a boy		-0.435**
		(0.194)
1st wife boy x 2nd wife boy		-0.021
		(0.270)
Observations	1047	799
R-squared	0.129	0.140
Mean/Mean control	1.011	1.061
Dependent variable: husband's tolerance of IPV	(1)	
Panel B: Monogamous husbands		
Husband's eldest is a boy	0.069	
	(0.053)	
Observations	3543	
R-squared	0.080	
Mean	0.834	

Note: The dependent variable is the score of the tolerance of IPV of the husband. I control for men's ethno-linguistic group, religion, age, age difference with the partner, age at first union, level of education, the region of residence, the DHS wealth score of the household, and the DHS survey wave. I compute robust standard errors. Significance level: * $p<0.10$, ** $p<0.05$, *** $p<0.01$. Sample: Men married to respondents in the main sample and who was surveyed for the men's survey. Source: DHS Burkina Faso 2003 and 2010.

2.5.3 Experience of IPV

As shown in table 2.11, I find that, among women married to a polygamous partner, having two first born daughters is negatively correlated with emotional violence (-30% compared to the mean), and that having three daughters among the first three children is also negatively correlated with experiencing physical violence (-44%). Given that the estimates yielded by the regressions are imprecise, I cannot rule out that that the sex of the first born child nor of the first three children have no impact (or large impact for that matter) on any of the type of violence studied. When I split the sample according to the rank of the spouses as in column (1) of table 2.12, I find that, for first spouses, having a first born boy is negatively correlated with experiencing a controlling behavior from the partner (-0.25% compared to the mean). The result is not robust however to interacting the sex of the first child of the first wife with the sex of the first child of the second wife. Panel B and column (3) of table 2.12 shows that being a second wife who has a boy rather than a girl is associated with less emotional violence, and column (5) suggests that for second wives, having a first wife whose first child is a boy is negatively correlated with experiencing physical IPV. As for women in a monogamous union, table 2.13 shows that having a first born son is positively correlated with the experience of emotional violence (+15% compared to the mean level), and with the experience of physical violence (+ one third compared to the mean). Columns (2) and (3) of table 2.13 suggests that having two or three first born daughters is negatively associated with experiencing a controlling behavior.

As with the measures of agency, I exploit the information available on the timing of the arrival of the first son after the first marriage of women; column (5) of panel A of table 2.14 suggests that the higher the probability to still not have a son at time t, the more likely women in a polygamous union are to experience physical IPV. Columns (1) and (2) of panel B of table 2.14 shows that among women in a monogamous union, a higher probability to still not have a son at time t is correlated with experiencing more controlling behavior and more emotional violence.

Table 2.11: Experience of IPV among women in a polygamous union

	index control			emotional violence			index physical ipv		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel: Women in a polygamous union									
Boy first born	-0.052			-0.010			0.056		
	(0.066)			(0.015)			(0.084)		
GG		0.070			-0.032*			-0.132	
		(0.090)			(0.017)			(0.095)	
GGG			0.096			-0.033			-0.216**
			(0.121)			(0.022)			(0.105)
Observations	2081	2016	1959	2081	2016	1959	2081	2016	1959
R-squared	0.109	0.110	0.105	0.100	0.104	0.103	0.093	0.096	0.095
Mean	0.985	0.986	0.975	0.110	0.110	0.110	0.492	0.497	0.491

Note: The dependent variables index control and index physical IPV are scores yielded by a PCA and rescaled to take positive values. The variable emotional violence is a dummy equal to 1 if the respondent ever experienced emotional violence, 0 otherwise. GG stands for the first two children of the respondent are girls. GGG stands for the first three children of the respondent are girls. I control for women's ethno-linguistic group, religion, age category, the age difference with the partner, the age at first union, the duration of the union, the total number of birth, the husband's education, the region of residence, the DHS wealth score of the household, and the DHS survey wave. I compute robust standard errors. Significance level: * $p<0.10$, ** $p<0.05$, *** $p<0.01$. Sample: Women aged 17-49 years old married to a polygamous partner. Source: DHS Burkina Faso 2010.

Table 2.12: Experience of IPV and sex of the first child - Polygamous spouses

	index control		**emotional violence**		**index physical ipv**	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: First wives						
1st wife's child is a boy	-0.231**	-0.214	0.002	0.021	-0.067	-0.297
	(0.115)	(0.171)	(0.024)	(0.040)	(0.145)	(0.206)
2nd wife's child is a boy	-0.169	-0.151	-0.030	-0.009	-0.025	-0.268
	(0.113)	(0.181)	(0.028)	(0.039)	(0.141)	(0.230)
1st wife boy x 2nd wife boy		-0.035		-0.041		0.484
		(0.222)		(0.054)		(0.305)
Observations	592	592	592	592	592	592
R-squared	0.132	0.132	0.154	0.155	0.105	0.109
Mean/Mean control	0.906	1.054	0.120	0.095	0.567	0.700
	index control		**emotional violence**		**index physical ipv**	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel B: Second wives						
1st wife's child is a boy	0.036	0.151	0.003	0.043	-0.268**	-0.039
	(0.126)	(0.136)	(0.026)	(0.042)	(0.128)	(0.175)
2nd wife's child is a boy	-0.004	0.111	-0.054*	-0.014	0.065	0.294
	(0.121)	(0.196)	(0.029)	(0.036)	(0.134)	(0.256)
1st wife boy x 2nd wife boy		-0.217		-0.075		-0.429
		(0.247)		(0.054)		(0.303)
Observations	588	588	588	588	588	588
R-squared	0.117	0.119	0.146	0.149	0.158	0.162
Mean/Mean control	1.097	1.073	0.109	0.114	0.482	0.593

Note: The dependent variables index control and index physical IPV are scores yielded by a PCA and rescaled to take positive values. The variable emotional violence is a dummy equal to 1 if the respondent ever experienced emotional violence, 0 otherwise. I control for women's ethno-linguistic group, religion, age category, the age difference with the partner, the age at first union, the duration of the union, the total number of birth, the husband's education, the region of residence, the DHS wealth score of the household, and the DHS survey wave. I compute robust standard errors. Significance level: * $p<0.10$, ** $p<0.05$, *** $p<0.01$. Sample: Panel A: women aged 17-49 years old, married to a polygamous partner and who declare that they are the first wife of their partner. Panel B: women aged 17-49 years old, married to a polygamous partner and who declare that they are the second wife of their partner. Source: DHS Burkina Faso 2010.

Table 2.13: Experience of IPV and sex of the first child - Monogamous

	index control			emotional violence			index physical ipv		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel: Women in a monogamous union									
Boy first born	0.038			0.014*			0.150***		
	(0.046)			(0.008)			(0.044)		
GG		-0.137**			-0.007			-0.018	
		(0.055)			(0.010)			(0.056)	
GGG			-0.142**			0.001			-0.025
			(0.072)			(0.018)			(0.091)
Observations	6061	5540	5278	6061	5540	5278	6061	5540	5278
R-squared	0.074	0.072	0.077	0.072	0.075	0.077	0.052	0.050	0.051
Mean	1.042	1.029	1.024	0.095	0.096	0.096	0.450	0.470	0.473

Note: The dependent variables index control and index physical IPV are scores yielded by a PCA and rescaled to take positive values. The variable emotional violence is a dummy equal to 1 if the respondent ever experienced emotional violence, 0 otherwise. GG stands for the first two children of the respondent are girls. GGG stands for the first three children of the respondent are girls. I control for women's ethno-linguistic group, religion, age category, the age difference with the partner, the age at first union, the duration of the union, the total number of birth, the husband's education, the region of residence, the DHS wealth score of the household, and the DHS survey wave. I compute robust standard errors. Significance level: * $p<0.10$, ** $p<0.05$, *** $p<0.01$. Sample: Women aged 17-49 years old, married to a monogamous partner. Source: DHS Burkina Faso 2010.

Table 2.14: Experience of IPV and probability not to have a son at time t

	index control		**emotional IPV**		**index physical IPV**	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Polygamous wives						
1st wife: time btw first union and first boy	0.008	-0.057	0.029	0.057	0.333*	0.430
	(0.14)	(0.20)	(0.03)	(0.04)	(0.18)	(0.29)
2nd wife	0.150*	0.078	-0.042**	-0.010	-0.215*	-0.108
	(0.09)	(0.17)	(0.02)	(0.04)	(0.12)	(0.17)
2nd wife: time btw first union and first boy		0.123		-0.055		-0.184
		(0.26)		(0.05)		(0.33)
Observations	2086	2086	2086	2086	2086	2086
R-squared	0.047	0.047	0.038	0.038	0.060	0.061
Unconditional Mean	0.988	0.988	0.097	0.097	0.448	0.448

	index control		**emotional IPV**		**index physical IPV**	
	all	first union	all	first union	all	first union
Panel B: Monogamous wives						
time btw first union and first boy	0.182**	0.207**	0.016	0.017	0.108	0.076
	(0.09)	(0.09)	(0.02)	(0.02)	(0.09)	(0.09)
Observations	6177	5641	6177	5641	6177	5641
R-squared	0.032	0.029	0.026	0.023	0.022	0.024
Unconditional Mean	1.069	1.049	0.084	0.082	0.375	0.373

Note: The probability to still not have a son at time t is the predicted survival function to knowing the event "the birth of the first son". It is obtained using a standard semiparametric Cox model stratified by the polygamous status of the respondent. Variables used to predict the survival function are a set of characteristics of the respondent that are determined at the time of the first union: region of residence, wealth of the household, religious denomination of the respondent, and ethno-linguistic group. Results presented in panel A and panel B are yielded by a LPM. In the regressions presented in the table, the controls included are women's ethno-linguistic group, religion, age category, the age difference with the partner, the age at first union, the duration of the union, the total number of birth, and the DHS survey wave. Sample of Panel A: Women aged 17-49 years old, married to a polygamous partner and who declare that they are the first wife or the second wife of their partner. Sample of Panel B: Women aged 17-49 years old, married to a monogamous partner. Source: DHS Burkina Faso 2010.

2.5.4 Summary of Results and Discussion

To summarize the results, I cannot provide a clear-cut nor causal answer to the question "how does the sex composition of children impact the mother's status, agency, and access to resources in the household". It seems that for first wives, having a first born son is positively correlated with having more say in the decision-making process. It suggests that first wives whose first child is a boy benefit from an advantage in decision-making compared to first wives whose first child is a girl, and to second wives, irrespective of the sex of their eldest child. When exploring whether the impact of the sex composition of the other spouse matters for women's own agency, I find that it may matter for first wives; being a first wive who has at least one boy among her first three children and whose junior spouse only has girls among her first three children, is positively correlated with making decision on own's own. The tests for anemia are probably underpowered, which does not allow to provide a definitive answer about the impact of the first born child on the mother's health.

I also find that being a second wife with two girls among the first two children is positively correlated with tolerating IPV compared to first wives, and to second wives whose first child is a boy. It suggests that second wives whose first two children are girls may incur a penalty in terms of tolerance of IPV compared to the other partners of a polygamous man. This result is yielded by a LPM among wives whose cowife also has two girls among her first two children, and thus may mix the impact of the sex composition of the first two children with unobserved characteristics of the households. The LPM also suggests that among second wives, having at least one boy is negatively correlated with the probability to be anemic compared to first wives with at least one boy among the first two children, and to second wives with two girls among the first two children. I also find evidence that for second spouses, having a first son at the very beginning of the union seem negatively correlated with the probablility to be anemic. Overall, the results are compatible with the suggestion that having a son matters for second wife as it is correlated with a better status in the household (proxied by tolerance of IPV) and with an increased access to resources proxied by the probability to be anemic.

I also find results compatible with the hypothesis that the sex composition of the children of a co-spouse may influence the other wife's agency and experience of IPV.

The descriptive results summarized here potentially point to complex effects of the sex compo-

sition of children on their mother's status, agency, and access to resources in the context of West Africa. They may notably be subject to dynamic changes, that are difficult to fully capture with the data used in the analysis. I indeed provide descriptive evidence that, among polygamous wives, the timing of the birth of a first son impacts the mother's tolerance of IPV, as the data suggests that as the probability to still not have a son at time t increases, the tolerance of IPV changes less among second wives than it does among first wives.

While the LPM model does not allow to conclude whether the sex of the first child impacts monogamous women's tolerance of IPV, I find that for monogamous women, the higher the probability to still not have a son at time t, the less they tolerate IPV. The estimation is underpowered to detect an impact on decision-making or anemia. I also find that being the monogamous mother of a first child who is a boy is correlated with more emotional and physical violence, and that having two girls or three girls among the first two or three children is associated with less controlling behavior. The study of the timing of the birth of the first son might nuance the results, as the probability to still not have a son at time t is positively correlated with experiencing controlling behavior among monogamous women. As for women married to a polygamous partner, I also find that having two or three girls among the first two or three children is associated with less IPV, but results suggest that, as for monogamous women, the higher the probability to still not have a son at times t among polygamous wives, the more likely they are to experience physical IPV.

If the results are taken at face value, they suggest that among women married to a polygamous partner, first wives who have a first born son seem to have more autonomy in the decision-making process than first wives whose first child is a girl, and than second wives regardless of the sex of their first child. As for second wives, the results may hint at the fact that there is a premium in having at least one son among the first two children in terms of actual access to resources; anemia-wise, second wives with at least one son seem to be better off than first wives, regardless of the gender of the first two children, and that second wives with only daughters among the first two children. As for the experience of IPV, results are compatible with the hypothesis that having a son remains key, as the longer the expected wait for the birth of the first son, the more women may experience physical IPV. But having a first born son may also expose women to a backlash effect, possibly in response to them claiming more resources in the household, or by making them subject to more controlling behavior as they are now in charge of the education of the eldest heir of the husband. This interpretation would be consistent with Donald *et al.* (2021) that shows that sole decision making by the wife is associated

with a 10 percentage point higher incidence of IPV based on an anlysis made in 12 Sub-Saharan countries, including Burkina Faso.

Even though some of the tests are underpowered to rule out a null or large impact of the gender composition of children on the outcomes of interest, the results may suggest that tolerance of IPV, which proxies women's status in the household, autonomy in decision-making, and anemia, which proxies women's actual access to resources, are not aligned in this study. It raises questions about the meaning that such questions take for respondents. Annan *et al.* (2021) notably shows using the DHS that in 23 Sub-Saharan countries, husbands and wives provide conflicting answers to who is the main decision-maker in the household, and that the fact that women declare being the main decision-maker when their husband does not, actually reflects a dynamic of power in the household. In a forthcoming working paper, they also suggest that husband and wife do not have the same understanding of what being the main decision-maker in the household means. These considerations suggest that there may be several layers of interpretations of the results presented here that the present study cannot disentangle. The difference in levels of tolerance of IPV between the two survey waves used in the paper, as shown by Figure 2.2, also fuels the concerns about the consistency of the measure across time. The interpretation of results on the tolerance of IPV may be made more complex by the fact that, even though women's ability to express disagreement with a restricitve gender norms speaks to a certain form of their agency,[15] the link with their status may be ambiguous. First, the question of status is related to the attitudes or behaviors that a community enhance by valuing them. As pointed out by Kabeer (1999), women may be faced with a trade-off between their ability to derive status within the family or community and their ability to make independent choices or express independent opinions that may reflect gender-incongruent attitudes or behaviors. Kabeer (1999) goes on underlining that it may be especially true in high-fertility contexts where there exists a preference for sons as women may be granted rights and privileges according to their number of children, and potentially their number of boys. In addition, attitudes towards IPV may well reflect the general opinion in a woman's community rather than her own (Novak (2016), Bellemare *et al.* (2015)), or some women may express support for a gender unequal norm because they may benefit from it, or because they know they would not suffer from it. For instance, a senior spouse condemning IPV if she were to be the victim of it, but condoning its use against a junior wife. Further research is necessary to assess whether the apparent disconnect between tolerance of IPV, decision-making in the household, anemia,

[15] The measure has been validated as a measure of intrinsic agency, that is defined in Miedema *et al.* (2018) as "the extent to which women's expression of gender attitudes reflect or reject normative beliefs", in the context of Egypt (Yount *et al.* (2016)) and East Africa (Miedema *et al.* (2018)) notably.

and experience of IPV results from the use of flawed tools whose meaning differ for researchers and respondents.

The intepretation of the results for women married to a monogamous partner remain a puzzle. The gender composition of the children of a monogamous mother seems to matter through the probability to still not have a son at time t; the higher it is, the less women tolerate IPV, and the higher the association with experiencing a controlling behavior from the partner. In addition, having a first born child who is a son is correlated with more emotional violence, and with more physical violence. The results could be a symptom that for monogamous having a son matters but once the first heir to the partner's lineage is born, women are more monitored and controlled by the partner. This interpretation is in line with Yount (2005) who concludes that married women in a situation of dependence on their husbands (dependence being defined as having a lower level of education than the husband and having several boys) are more exposed to IPV. Yount (2005) provides descriptive evidence that, even in contexts that display a preference for sons in their fertility behavior as evidenced by Rossi and Rouanet (2015) in Egypt it cannot be ruled out that having boys may enhance their mother's exposure to IPV. The results among monogamous wives could also be driven by a differential tendency to report IPV between monogamous mothers of a first born girl and monogamous mothers of a first born boy.

Subsequent versions of this paper would benefit from the addition of a new wave of the DHS for Burkina Faso when it becomes available to improve the statistical power of the study.

Table 2.15: Summary of main results

		Score Tolerance IPV		Score Decision-making		Anemia		Tolerance of IPV of husband		Experience IPV	
sex composition of children	*Reference*	Mean/Mean control	Δ	Mean/Mean control	Δ	Mean/Mean control	Δ	Mean/Mean control	Δ	Mean/Mean control	Δ
boy first born polygamous wives											
boy first born monogamous wives										0.09 0.45	emotional ipv: +15\% physical ipv: +33\%
GG polygamous wives										0.11	emotional ipv: -0.32\%
GG monogamous wives										1.03	control : -13\%
GGG polygamous wives										0.49	physical ipv: -44\%
GGG monogamous wives										1.02	control: -14\%
boy first born x 1st wife	**girl first born x 1st wife**			0.61	(FE) +14\%					0.91	control: -25\%
boy first born x 2nd wife								1.06	(LPM) -41\%	0.11 0.48	emotional ipv: -50\% physical ipv: -56\%
GG x 1st wife											
GG x 2nd wife	**GB, BG, or BB x 1st wife**	1.04	(LPM not contrib. to FE) +12\%								
GB, BG or BB x 2nd wife	**GB, BG, or BB x 1st wife**					0.54	(LPM contrib to FE) -26\%				
GB, BG or BB x 2nd wife	**GB, BG, or BB x 1st wife**					0.55	(LPM not contrib to FE) -10\%				
being a first wife with 2nd wife GGG	**being a first wife with 2nd wife with at least 1 boy out of 3 children**			0.54	(LPM) +33\%						
probability of still not having a son time t											
2nd wife who has a boy at the onset of the union	unconditional mean					0.53	-30\%			0.45	physical ipv: -53\%
Monogamous wives	unconditional mean	0.87	-10\%							1.07	control: -17\%

Note: In the sex composition of children, GG stands for the first two children of the respondent are girls, and GGG stands for the first three children of the respondent are girls. GB stands for the first two children of a respondent are a girl and a boy, and BB stands for the first two children of a respondent are two boys.

2.6 Conclusion

This paper addresses whether the sex composition of children, and whether the sex of the first born child in particular, decreases the mother's tolerance of IPV, autonomy in decision-making, access to resources proxied by anemia, and whether it shields her from experiencing IPV. To answer this question, I use the DHS for Burkina Faso and exploit the high level of polygamy to apply a husband fixed-effect as my preferred specification, as it deals with what would be a main omitted variable otherwise: the characteristics of the husband. As it is demanding on the data, I also use a linear probability model, even though it does not allow me to claim causality.

I am not able to provide conclusive evidence of the impact of the sex composition of children on their mother's agency or experience of IPV. The data however seem compatible with the following patterns: first wives whose first child is a boy seem to benefit from more autonomy in the decision-making process, being the second wife and having two daughters among the first two children seem correlated with incurring a penalty in terms of tolerance of IPV. Using the predicted probability to still not have a son at time t, I also find that for monogamous women, having a son is associated with less tolerance of IPV, and with more reported controlling behavior by the partner. For polygamous women, the predicted probability to still not have a son at time t is positively correlated with experiencing more physical IPV, and the results yielded by a LPM sugests that when a first spouse has a first born son, it may decrease experience of IPV for herself but also for the second spouse. As for second wives, having a first born son is negatively correlated with experiencing emotional violence. Interestingly, the impact of having a first born son estimated with a LPM yields a positive association with IPV for women married to a monogamous partner.

The results, which are mainly descriptive, have to be interpreted with caution. Tolerance of IPV, autonomy in decision-making, access to resources, and to a certain extent experience of IPV are meant to reflect different facets of women's autonomy and empowerment. This paper underlines that, depending on the rank of the spouse, they seem to move in different directions in response to the sex composition of children, which invites to a reflection on the consistency and the quality of the tools currently used in large household surveys to proxy women's autonomy and empowerment. Reported experience of IPV is specific in its own right as it is a particularly sensitive topic that is prone to declaration bias. Enriching the way women are asked about IPV using experimental approach may prove fruitful to circumvent the issue of declaration bias.

Despite the limitations, this paper contributes to the body of literature on the impact of family composition on domestic violence and more broadly to the literature on domestic violence in developing countries. It also provides food for thought on the channels of expression of a preference for boys in sub-Saharan Africa.

Appendix

Table A-2.1: Correlation of tolerance of IPV among polygamous women and husbands

	Wife' s tolerance				
	goes out w/o telling	neglects children	argues with husband	refuses sex	burns the food
husband's tolerance					
goes out w/o telling	0.0492*	0.0544*	0.0511*	0.0530*	0.0714*
neglects children	0.0154	0.0333	0.0174	0.0288	0.0464*
argues with husband	0.0136	0.0415*	0.0644*	0.0196	0.0263
refuses sex	0.0049	-0.0094	0.0077	0.01	0.0206
burns the food	0.0101	0.0292	0.0258	0.0186	0.0261

Note: Sample: Women aged 17-49 years old, married to a polygamous partner with exactly two wives, and whose husband was surveyed in the men survey. Source: DHS Burkina Faso 2003 and 2010. * denotes significance level of 5% or less. Adjustment for MHT by Bonferroni method.

Table A-2.2: Correlation matrix of tolerance of IPV among monogamous women and husbands

	Wife' s tolerance				
	goes out w/o telling	neglects children	argues with husband	refuses sex	burns the food
husband's tolerance					
goes out w/o telling	0.0655*	0.0527*	0.0549*	0.0662*	0.0615*
neglects children	0.0429*	0.0273	0.0395*	0.0509*	0.0602*
argues with husband	0.0495*	0.0495*	0.0639*	0.0612*	0.0490*
refuses sex	0.0603*	0.0370*	0.0521*	0.0783*	0.0746*
burns the food	0.0485*	0.0494*	0.0419*	0.0749*	0.0571*

Note: Sample: Women aged 17-49 years old, married to a monogamous partner, and whose husband was surveyed in the men survey. Source: DHS Burkina Faso 2003 and 2010. * denotes significance level of 5% or less. Adjustment for MHT by Bonferroni method.

Table A-2.3: Test of birth parities of boys and girls

Variables	Observations	Mean	Std. Err.	Std. Dev.	Bounds of 95% Conf. Interval		pval: Ha: mean(diff) $\neq 0$	pval: Ha: mean(diff) < 0
Mean rank of girls	15,025	3.16	0.01	1.44	3.14	3.18		
Mean rank of boys	15,025	3.17	0.01	1.44	3.15	3.2		
difference	15,025	-0.01	0.01	1.58	-0.04	0.01	0.3	0.15

Note: Sample: Women aged 17-49 years old who had at least one birth. Source: DHS Burkina Faso 2003 and 2010.

Table A-2.4: Sex ratio at birth per age category of respondent

Age category of the respondent	sex ratio
17-24	105.6261
25-29	103.9301
30-34	105.9537
35-39	103.6346
40-44	105.262
45-49	104.4772
Total	105.1649

Note: Sample: Women aged 17-49 years old. Source: DHS Burkina Faso 2003 and 2010.

Table A-2.5: Balancing tests 2nd wives in first union vs. not in first union

	Second wives				Third wives			
Variables	Not in first union	In first union	difference	N	Not in first union	In first union	difference	N
age	34.935	30.838	-4.108***	4.477	36.103	30.676	-5.195***	1.597
	[7.908]	[8.462]	(0.312)		[7.126]	[8.345]	(0.546)	
No formal education	0.905	0.923	0.014	4.477	0.918	0.932	0.017	1.597
	[0.293]	[0.267]	(0.01)		[0.275]	[0.252]	(0.017)	
some primary education	0.081	0.057	-0.022**	4.477	0.079	0.053	-0.03*	1.597
	[0.273]	[0.232]	(0.009)		[0.271]	[0.224]	(0.016)	
some secondary education	0.014	0.02	0.007	4.477	0.002	0.015	0.013*	1.597
	[0.116]	[0.14]	(0.005)		[0.05]	[0.121]	(0.007)	
age at first cohabitation	17.274	17.422	-0.02	4.477	17.385	17.48	-0.05	1.597
	[3.037]	[2.66]	(0.101)		[2.462]	[2.427]	(0.163)	
age at first birth	19.3	18.952	-0.432***	4.241	19.308	19.13	-0.28	1.52
	[3.516]	[3.148]	(0.123)		[3.509]	[2.896]	(0.209)	
total number of births	4.879	4.053	-0.779***	4.477	4.875	3.979	-0.874***	1.597
	[2.716]	[2.643]	(0.099)		[2.49]	[2.541]	(0.171)	
age of partner	47.72	48.391	0.324	4.268	53.35	57.169	2.871***	1.509
	[11.252]	[12.289]	(0.459)		[11.437]	[12.026]	(0.81)	
sex of the first birth	0.541	0.511	-0.023	4.241	0.446	0.528	0.072**	1.52
	[0.499]	[0.5]	(0.019)		[0.498]	[0.499]	(0.035)	
first child survived at least until 1	0.859	0.885	0.02	4.241	0.829	0.892	0.065***	1.52
	[0.348]	[0.319]	(0.012)		[0.378]	[0.311]	(0.022)	
Partner has no formal education	0.879	0.915	0.033***	4.477	0.928	0.96	0.026*	1.597
	[0.326]	[0.279]	(0.01)		[0.259]	[0.197]	(0.014)	
partner has some primary education	0.075	0.058	-0.015*	4.477	0.06	0.034	-0.024*	1.597
	[0.263]	[0.234]	(0.009)		[0.239]	[0.181]	(0.013)	
partner has some secondary education	0.029	0.016	-0.013***	4.477	0.003	0.001	-0.002	1.597
	[0.167]	[0.124]	(0.005)		[0.053]	[0.037]	(0.003)	
DHS wealth quintiles								
poorest	0.207	0.178	0.000	4.477	0.151	0.13	0.000	1.597
	[0.405]	[0.383]	(0.000)		[0.359]	[0.337]	(0.000)	
poorer	0.238	0.211	0.000	4.477	0.262	0.164	0.000	1.597
	[0.426]	[0.408]	(0.000)		[0.441]	[0.37]	(0.000)	
middle	0.245	0.265	0.000	4.477	0.25	0.301	0.000	1.597
	[0.43]	[0.441]	(0.000)		[0.434]	[0.459]	(0.000)	
richer	0.184	0.223	0.000	4.477	0.22	0.285	0.000	1.597
	[0.388]	[0.416]	(0.000)		[0.415]	[0.452]	(0.000)	
richest	0.126	0.123	0.000	4.477	0.117	0.12	0.000	1.597
	[0.332]	[0.329]	(0.000)		[0.322]	[0.325]	(0.000)	
muslim	0.428	0.447	0.000	4.477	0.391	0.397	0.000	1.597
	[0.495]	[0.497]	(0.000)		[0.489]	[0.489]	(0.000)	
mossi	0.497	0.593	0.000	4.477	0.575	0.677	0.000	1.597
	[0.5]	[0.491]	(0.000)		[0.495]	[0.468]	(0.000)	
justifies wife beating for common offense	0.639	0.589	-0.021	4.477	0.73	0.636	-0.093***	1.597
	[0.481]	[0.492]	(0.018)		[0.445]	[0.481]	(0.031)	
justifies wife beating for severe offense	0.414	0.361	-0.012	4.477	0.456	0.402	-0.046	1.597
	[0.493]	[0.48]	(0.017)		[0.499]	[0.49]	(0.032)	
Partner justifies wife beating for								
going out w/o telling him	0.187	0.197	0.003	1.362	0.236	0.226	0.003	323
	[0.39]	[0.398]	(0.026)		[0.428]	[0.419]	(0.053)	
neglects children	0.203	0.229	0.028	1.361	0.214	0.195	-0.028	324
	[0.403]	[0.42]	(0.028)		[0.413]	[0.397]	(0.056)	
argues with husband	0.214	0.237	0.052*	1.353	0.341	0.235	-0.064	322
	[0.411]	[0.426]	(0.028)		[0.477]	[0.425]	(0.055)	
refuses sex	0.138	0.139	0.000	1.356	0.08	0.141	0.036	323
	[0.346]	[0.346]	(0.023)		[0.273]	[0.349]	(0.044)	
burns the food	0.112	0.093	-0.014	1.36	0.033	0.063	0.045	321
	[0.316]	[0.29]	(0.02)		[0.179]	[0.243]	(0.033)	
anemic	0.554	0.531	-0.023	1.937	0.469	0.549	0.061	629
	[0.498]	[0.499]	(0.029)		[0.501]	[0.498]	(0.053)	
sole decision making score	0.739	0.552	-0.175***	4.477	0.758	0.520	-0.079	1.597
	[1.331]	[1.092]	(0.0424)		[1.383]	[1.077]	(0.076)	

Note:Significance level: * p<0.10, ** p<0.05, *** p<0.01. Sample: Women aged 17-49 years old, married to a polygamous partner and who declare that they are the second wife or the third wife of their partner. Source: DHS Burkina Faso 2003 and 2010.

Table A-2.6: Women's agency and sex of the first born - Spouses of rank 1, 2 or 3

	score tolerance IPV			score decision making			anemia		
	FE	LPM contrib. to id	LPM not contributing	FE	LPM contrib. to id	LPM not contributing	FE	LPM contrib. to id	LPM not contributing
Panel: 2 or 3 wo-wives ranked 1, 2 or 3									
Boy first born	-0.012	0.003	0.003	0.121***	0.066	0.066	-0.000	0.040	0.028
	(0.05)	(0.06)	(0.06)	(0.04)	(0.06)	(0.06)	(0.05)	(0.04)	(0.04)
2nd wife	-0.050	0.032	0.032	0.058	0.037	0.037	-0.039	0.005	-0.023
	(0.05)	(0.07)	(0.07)	(0.05)	(0.06)	(0.06)	(0.05)	(0.05)	(0.04)
3rd wife	-0.064	0.030	0.030	0.079	0.106	0.106	-0.049	-0.053	-0.032
	(0.08)	(0.10)	(0.10)	(0.10)	(0.11)	(0.11)	(0.08)	(0.07)	(0.06)
Boy first x 2nd wife	0.000	-0.061	-0.061	-0.136**	-0.033	-0.033	0.028	-0.064	-0.035
	(0.06)	(0.10)	(0.10)	(0.07)	(0.10)	(0.10)	(0.06)	(0.06)	(0.05)
Boy first x 3rd wife	0.045	0.098	0.098	-0.156	-0.166	-0.166	0.068	0.101	-0.099
	(0.09)	(0.13)	(0.13)	(0.11)	(0.14)	(0.14)	(0.09)	(0.09)	(0.08)
Observations	8290	3070	3070	8290	3070	3070	3336	1379	2262
Contributing to id	3275			3275			3275		
R-squared	0.009	0.077	0.077	0.016	0.013	0.013	0.022	0.033	0.019
Mean Control	1.053		1.072	0.577		0.494	0.521		0.522
Husbands	5623			5623			5623		
Contributing to id	1507			1507			1507		
pval test: boy first + boy first x 2nd wife = 0	0.796			0.747					
pval test: boy first + boy first x 3rd wife = 0	0.684			0.729					

Note: The dependent variables are a score of justifying wife beating, a score of sole decision-making in the household, and the probability to have anemia. In columns 1, 4, and 7, I report the results yielded by the husband fixed effect (FE) estimation, and in columns 2, 5, and 8, I report the results yielded by the linear probability model (LPM) on the subsample of respondents who contribute to identifying the husband FE model. In columns 3, 6, and 9, I report the estimates obtained by using a LPM on the subsample of wives who do not participate to the identification of the husband fixed-effect. In the husband FE, I control for women's ethno-linguistic group, religion, age category, the age difference with the partner, the age at first union, the duration of the union, the total number of birth, and the DHS survey wave. For anemia, I also control whether women are pregnant at the time of the survey. I cluster standard errors at the husband level. In the LPM, I also control for the husband's education, the region of residence, and the DHS wealth score of the household. I compute robust standard errors. Sample: Women aged 17-49 years old, married to a polygamous partner and who declare that they are the first, second or third wife of their partner. Source: DHS Burkina Faso 2003 and 2010.

Table A-2.7: Summary statistics by sex of the first child

	All	Boy	Girl	Diff	p-value
Age	32.02	32.05	31.99	0.06	0.589
Respondent's education					
no formal education	0.84	0.84	0.84	0.01	0.282
primary	0.11	0.1	0.11	0	0.588
secondary and more	0.06	0.05	0.06	0	0.319
age at first cohabitation	17.46	17.44	17.48	-0.04	0.255
being in first union	0.88	0.88	0.88	0	0.653
age at first birth	19.03	19.03	19.03	0	0.975
total number of births	4.36	4.35	4.38	-0.03	0.431
first child survived at least until 1	0.89	0.88	0.9	-0.02	0
total birth alive	3.56	3.54	3.59	-0.05	0.097
boys alive	1.82	2.22	1.43	0.79	0
girls alive	1.79	1.37	2.2	-0.83	0
in a polygamous union	0.46	0.46	0.46	0	0.514
Rank of the spouse					
1st spouse	0.39	0.39	0.39	0	0.847
2nd spouse	0.45	0.45	0.45	0	0.897
3rd spouse	0.12	0.12	0.12	0	0.943
4th spouse and +	0.04	0.04	0.04	0	0.35
Age of the partner	43.37	43.41	43.34	0.07	0.69
Age difference with the partner					
more than 10 years	0.48	0.48	0.48	0	0.85
5 to 9 years	0.32	0.32	0.32	0	0.482
less than 5 years	0.2	0.2	0.21	0	0.563
Education of the partner					
no formal education	0.81	0.81	0.81	0	0.45
primary	0.1	0.1	0.1	0	0.616
secondary and +	0.07	0.07	0.07	0	0.758
DHS wealth score					
poorest	0.18	0.18	0.18	0	0.389
poorer	0.2	0.2	0.2	0	0.666
middle	0.23	0.22	0.23	0	0.743
richer	0.2	0.2	0.2	0	0.677
richest	0.2	0.2	0.2	0	0.636
muslim	0.46	0.46	0.45	0.01	0.187
mossi	0.52	0.51	0.52	-0.01	0.032
tolerance IPV common offense	0.57	0.56	0.57	-0.01	0.4
tolerance IPV severe offense	0.33	0.33	0.33	0	0.546
index tolerance husband	0.29	0.29	0.3	-0.01	0.576
anemic	0.51	0.5	0.52	-0.02	0.085
score sole decider	0.66	0.68	0.65	0.03	0.117
index controling behavior	1.02	1.01	1.03	-0.02	0.528
emotional violence	0.1	0.1	0.1	0	0.823
index physical violence	0.46	0.49	0.42	0.07	0.056
Observations	22821	10820	10155		

Note: Sample: Women aged 17-49 years old. Source: DHS Burkina Faso 2003 and 2010.

Table A-2.8: Summary statistics by sex of the first child among polygamous

All	Boy	Girl	Diff	p-value	
Age	33.63	33.65	33.61	0.04	0.834
Respondent's education					
no formal education	0.92	0.92	0.92	0	0.66
primary	0.07	0.07	0.06	0	0.596
secondary and more	0.02	0.02	0.02	0	0.937
age at first cohabitation	17.27	17.25	17.28	-0.03	0.637
being in first union	0.84	0.84	0.85	0	0.757
age at first birth	18.92	18.92	18.91	0	0.956
total number of births	4.94	4.88	4.99	-0.1	0.1
first child survived at least until 1	0.88	0.87	0.9	-0.03	0
total birth alive	3.93	3.88	3.99	-0.1	0.037
boys alive	1.99	2.36	1.62	0.73	0
girls alive	1.99	1.58	2.39	-0.82	0
Rank of the spouse					
1st spouse	0.46	0.46	0.46	0	0.957
2nd spouse	0.54	0.54	0.54	0	0.957
Age of the partner	46.67	46.55	46.78	-0.23	0.415
Age difference with the partner					
more than 10 years	0.59	0.58	0.6	-0.02	0.05
5 to 9 years	0.26	0.27	0.25	0.02	0.086
less than 5 years	0.15	0.16	0.15	0.01	0.554
Education of the partner					
no formal education	0.89	0.88	0.89	-0.01	0.174
primary	0.08	0.08	0.07	0.01	0.141
secondary and +	0.02	0.03	0.02	0	0.172
DHS wealth score					
poorest	0.19	0.19	0.19	0	0.811
poorer	0.22	0.21	0.22	-0.01	0.371
middle	0.26	0.25	0.26	-0.01	0.351
richer	0.21	0.21	0.21	0	0.994
richest	0.14	0.14	0.13	0.02	0.046
muslim	0.46	0.46	0.46	0.01	0.485
mossi	0.53	0.52	0.55	-0.02	0.045
tolerance IPV common offense	0.6	0.59	0.6	-0.01	0.635
tolerance IPV severe offense	0.36	0.36	0.36	-0.01	0.656
index tolerance husband	0.34	0.31	0.36	-0.05	0.182
anemic	0.52	0.51	0.53	-0.02	0.23
score sole decider	0.63	0.65	0.6	0.05	0.1
index controling behavior	0.98	0.97	0.99	-0.02	0.727
emotional violence	0.11	0.11	0.11	-0.01	0.642
index physical violence	0.5	0.49	0.5	-0.01	0.934
Observations	7150	3553	3323		

Note: Sample: Women aged 17-49 years old, and married to a polygamous partner. Source: DHS Burkina Faso 2003 and 2010.

Table A-2.9: Summary statistics by sex of the first child among monogamous

	All	Boy	Girl	Diff	p-value
Age	30.4	30.41	30.39	0.02	0.9
Respondent's education					
no formal education	0.78	0.78	0.77	0.01	0.184
primary	0.14	0.14	0.14	0	0.472
secondary and more	0.09	0.08	0.09	-0.01	0.277
age at first cohabitation	17.59	17.55	17.63	-0.08	0.129
being in first union	0.91	0.91	0.91	0	0.971
age at first birth	19.07	19.05	19.09	-0.04	0.504
total number of births	3.9	3.9	3.89	0.01	0.851
first child survived at least until 1	0.9	0.9	0.91	-0.01	0.023
total birth alive	3.26	3.25	3.28	-0.03	0.437
boys alive	1.68	2.08	1.27	0.81	0
girls alive	1.63	1.22	2.05	-0.84	0
Age of the partner	38.85	38.93	38.77	0.16	0.46
Age difference with the partner					
more than 10 years	0.34	0.34	0.33	0.01	0.393
5 to 9 years	0.4	0.4	0.4	0	0.946
less than 5 years	0.26	0.26	0.27	-0.01	0.399
Education of the partner					
no formal education	0.74	0.74	0.73	0.01	0.164
primary	0.14	0.13	0.14	-0.01	0.238
secondary and +	0.11	0.11	0.11	0	0.789
DHS wealth score					
poorest	0.19	0.19	0.18	0.01	0.308
poorer	0.19	0.19	0.19	0.01	0.455
middle	0.19	0.2	0.19	0	0.913
richer	0.18	0.18	0.18	0	0.875
richest	0.24	0.24	0.25	-0.01	0.116
muslim	0.47	0.47	0.46	0.01	0.292
mossi	0.49	0.48	0.5	-0.02	0.066
tolerance IPV common offense	0.52	0.51	0.52	-0.01	0.544
tolerance IPV severe offense	0.29	0.28	0.29	-0.01	0.418
index tolerance husband	0.29	0.3	0.29	0.01	0.719
anemic	0.49	0.48	0.5	-0.02	0.133
score sole decider	0.61	0.62	0.61	0.01	0.72
index controling behavior	1.04	1.04	1.04	0	0.922
emotional violence	0.1	0.1	0.09	0.01	0.471
index physical violence	0.45	0.5	0.41	0.09	0.031
Observations	11698	5617	5327		

Note: Sample: Women aged 17-49 years old, and married to a monogamous partner. Source: DHS Burkina Faso 2003 and 2010.

Table A-2.10: Being in first union and sex of the first two children

dependent variable *probability to be in first union at time of survey*	**All** (1)	**In a monogamous union** (2)	**In a polygamous union** (3)	**In a polygamous union** (4)
First two children are girls	-0.002	-0.004	-0.004	0.007
	(0.01)	(0.01)	(0.01)	(0.01)
2nd spouse			-0.198***	-0.194***
			(0.01)	(0.01)
3rd spouse			-0.191***	-0.185***
			(0.01)	(0.02)
GG x 2nd spouse				-0.019
				(0.02)
GG x 3rd spouse				-0.023
				(0.03)
Observations	17885	9435	8447	8447
Mean	0.877	0.908	0.843	0.843
pval test: boy 2nd wife = girl 2nd wife				0.520
pval test: boy 3rd wife = girl 3rd wife				0.546

Note: The dependent variable is being in first union. GG stands for the first two children of the respondent are girls. Controls: age, education level of the respondent, education level of the partner, difference in education between partners, age difference between partners, age at first cohabitation, total number of births, DHS wealth quintile, region of residence, being a muslim, being a mossi, DHS survey wave. Significance level: * p<0.10, ** p<0.05, *** p<0.01. Robust standard errors computed. Sample: Women aged 17-49 years old. Source: DHS Burkina Faso 2003 and 2010.

Table A-2.11: Being in first union and sex of the first three children

dependent variable *probability to be in first union at time of survey*	**All** (1)	**In a monogamous union** (2)	**In a polygamous union** (3)	**In a polygamous union** (4)
first 3 children are girls	-0.006	-0.015	-0.006	0.015
	(0.01)	(0.01)	(0.01)	(0.02)
2nd spouse			-0.195***	-0.192***
			(0.01)	(0.01)
3rd spouse			-0.193***	-0.189***
			(0.01)	(0.02)
GGG x 2nd spouse				-0.037
				(0.03)
GGG x 3rd spouse				-0.036
				(0.04)
Observations	17202	8989	8210	8210
Mean	0.876	0.907	0.843	0.843
pval test: boy 2nd wife = girl 2nd wife				0.403
pval test: boy 3rd wife = girl 3rd wife				0.595

Note:The dependent variable is being in first union. GGG stands for the first three children of the respondent are girls. Controls: age, education level of the respondent, education level of the partner, difference in education between partners, age difference between partners, age at first cohabitation, total number of births, DHS wealth quintile, region of residence, being a muslim, being a mossi, DHS survey wave. Significance level: * p<0.10, ** p<0.05, *** p<0.01. Robust standard errors computed. Sample: Women aged 17-49 years old. Source: DHS Burkina Faso 2003 and 2010.

Table A-2.12: Women's agency and sex of the first born child among women in a polygamous union

	score tolerance IPV			score decision making			anemia		
	FE	LPM contrib. to id	LPM not contributing	FE	LPM contrib. to id	LPM not contributing	FE	LPM contrib. to id	LPM not contributing
Panel: Two cowives of rank 1 and 2									
Boy first born	0.001	0.005	0.006	0.022	0.017	0.061	0.009	0.009	0.001
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.03)
Observations	6476	2047	4429	6476	2047	4429	2879	938	1941
Contributing to id	2047			2047					
R-squared	0.009	0.083	0.106	0.022	0.013	0.031	0.028	0.028	0.020
Mean	1.031			0.607					
Husbands	4454	1049	3405	4454	1049	3405			
Contributing to id	1049			1049					

height	score tolerance IPV			score decision making			anemia		
	FE	LPM contrib. to id	LPM not contributing	FE	LPM contrib. to id	LPM not contributing	FE	LPM contrib. to id	LPM not contributing
Panel: Two or three cowives of rank 1, 2 or 3									
Boy first born	-0.007	-0.013	0.028	0.040	0.032	0.046	0.021	0.022	0.000
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.02)
Observations	8290	3070	5220	8290	3070	5220	3641	1379	2262
Contributing to id	3070			3070					
R-squared	0.009	0.076	0.112	0.014	0.013	0.028	0.018	0.030	0.018
Mean	1.050			0.586					
Husbands	5283	1459	3824	5283	1459	3824			
Contributing to id	1459			1459					

Note: The dependent variables are a score of justifying wife beating, a score of sole decision-making in the household, and the probability to have anemia. In columns 1, 4, and 7, I report the results yielded by the husband fixed effect (FE) estimation, and in columns 2, 5, and 8, I report the results yielded by the linear probability model (LPM) on the subsample of respondents who contribute to identifying the husband FE model. In columns 3, 6, and 9, I report the estimates obtained by using a LPM on the subsample of wives who do not participate to the identification of the husband fixed-effect. In the husband FE, I control for women's ethno-linguistic group, religion, age category, the age difference with the partner, the age at first union, the duration of the union, the total number of birth, and the DHS survey wave. For anemia, I also control whether women are pregnant at the time of the survey. I cluster standard errors at the husband level. In the LPM, I also control for the husband's education, the region of residence, and the DHS wealth score of the household. I compute robust standard errors. Significance level: * p<0.10, ** p<0.05, *** p<0.01. Sample: Women aged 17-49 years old, married to a polygamous partner. Source: DHS Burkina Faso 2003 and 2010.

Table A-2.13: Women's agency and sex of the first born child among monogamous women

	score tolerance IPV (1)	**score decision making** (2)	**anemia** (3)
Panel: Two cowives of rank 1 and 2			
Boy first born	-0.011	0.005	-0.011
	(0.02)	(0.03)	(0.02)
Observations	10371	10371	4466
R-squared	0.181	0.059	0.043
Mean			

Note: The dependent variables are a score of justifying wife beating, a score of sole decision-making in the household, and the probability to have anemia. I control for women's ethno-linguistic group, religion, age category, the age difference with the partner, the age at first union, the duration of the union, the total number of birth, the DHS survey wave, whether the respondent is in first union, the education of the husband, the difference in education with the partner, the DHS wealth quintile fo the household, and the region of residence. For anemia, I also control whether women are pregnant at the time of the survey. I compute robust standard errors. Significance level: * p<0.10, ** p<0.05, *** p<0.01. Sample: Women aged 17-49 years old, married to a monogamous partner. Source: DHS Burkina Faso 2003 and 2010.

Table A-2.14: Women's agency and sex of the first three children

	score tolerance IPV			score decision making			anemia		
	FE	LPM	LPM no contrib	FE	LPM	LPM no contrib	FE	LPM	LPM no contrib
Panel: Two cowives of rank 1 and 2									
GGG	-0.061	-0.050	-0.100	0.001	-0.156	0.063	-0.005	-0.143	0.076
	(0.09)	(0.11)	(0.09)	(0.07)	(0.12)	(0.13)	(0.08)	(0.09)	(0.08)
2nd wife	-0.061	0.036	-0.001	-0.022	-0.166	0.004	-0.072*	-0.196*	-0.063**
	(0.05)	(0.13)	(0.03)	(0.05)	(0.13)	(0.04)	(0.04)	(0.10)	(0.03)
GGG x 2nd wife	0.134	0.130	0.250**	-0.083	0.301	-0.001	0.012	0.310**	0.091
	(0.13)	(0.18)	(0.12)	(0.13)	(0.19)	(0.16)	(0.12)	(0.12)	(0.10)
Observations	6061	710	5351	6061	710	5351	2682	313	2369
Contributing to id	710			710					
R-squared	0.009	0.253	0.129	0.023	0.089	0.045	0.030	0.148	0.032
Mean	1.019		1.036	0.641		0.621	0.525	0.508	0.527
Husbands	4310	365	3945	4310	365	3945			
Contributing to id	365			365					
pval test: GGG + GGG x 2nd wife = 0	0.481			0.451				0.071	0.014
pval test: GGG + 2nd wife + GGG x 2nd wife = 0								0.794	0.135

Note: The dependent variables are a score of justifying wife beating, a score of sole decision-making in the household, and the probability to have anemia. In columns 1, 4, and 7, I report the results yielded by the husband fixed effect (FE) estimation, and in columns 2, 5, and 8, I report the results yielded by the linear probability model (LPM) on the subsample of respondents who contribute to identifying the husband FE model. In columns 3, 6, and 9, I report the estimates obtained by using a LPM on the subsample of wives who do not participate to the identification of the husband fixed-effect. In the husband FE, I control for women's ethno-linguistic group, religion, age category, the age difference with the partner, the age at first union, the duration of the union, the total number of birth, and the DHS survey wave. For anemia, I also control whether women are pregnant at the time of the survey. I cluster standard errors at the husband level. In the LPM, I also control for the husband's education, the region of residence, and the DHS wealth score of the household. I compute robust standard errors. GGG stands for the first three children of the respondent are girls. Sample: Women aged 17-49 years old, married to a polygamous partner and who declare that they are the second wife or the third wife of their partner. Source: DHS Burkina Faso 2003 and 2010.

Table A-2.15: Women's agency and sex of the first child of the cospouse

	index tolerance		**index dexcision-making**		**anemia**	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: First wives						
1st wife's child is a boy	0.013	-0.011	-0.001	-0.023	0.034	0.026
	(0.050)	(0.067)	(0.053)	(0.071)	(0.034)	(0.050)
2nd wife's child is a boy	-0.006	-0.029	-0.014	-0.034	-0.006	-0.013
	(0.048)	(0.068)	(0.051)	(0.076)	(0.037)	(0.052)
1st wife boy x 2nd wife boy		0.046		0.041		0.015
		(0.095)		(0.104)		(0.069)
Observations	2032	2032	2032	2032	939	939
R-squared	0.086	0.086	0.043	0.043	0.057	0.057
Mean/Mean Control	1.014	1.046	0.554	0.546	0.537	0.534
Panel B: Second wives						
	index tolerance		**index decision-making**		**anemia**	
	(1)	(2)	(3)	(4)	(5)	(6)
1st wife's child is a boy	0.012	0.043	-0.073	-0.033	0.015	0.054
	(0.051)	(0.076)	(0.047)	(0.069)	(0.040)	(0.055)
2nd wife's child is a boy	0.009	0.039	-0.025	0.013	0.004	0.043
	(0.048)	(0.072)	(0.048)	(0.069)	(0.036)	(0.055)
1st wife boy x 2nd wife boy		-0.059		-0.074		-0.076
		(0.103)		(0.091)		(0.077)
Observations	2032	2032	2032	2032	941	941
R-squared	0.106	0.106	0.073	0.074	0.050	0.052
Mean/Mean control	1.002	0.997	0.509	0.502	0.522	0.498

Note: The dependent variables are a score of justifying wife beating, a score of sole decision-making in the household, and the probability to have anemia. I control for women's ethno-linguistic group, religion, age category, the age difference with the partner, the age at first union, the duration of the union, the total number of birth, the DHS survey wave, the husband's education, the region of residence, and the DHS wealth score of the household. For anemia, I also control whether women are pregnant at the time of the survey. I compute robust standard errors. Significance level: * p<0.10, ** p<0.05, *** p<0.01. Sample: Women aged 17-49 years old, married to a polygamous partner and who declare that they are the first wife or the second wife of their partner. Source: DHS Burkina Faso 2003 and 2010.

Table A-2.16: Women's agency and sex of the first two children of the cospouse

	index tolerance		**index decision-making**		**anemia**	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: First wives						
1st wife's first two children are girls	-0.088	-0.058	0.048	0.033	-0.068	-0.078
	(0.061)	(0.068)	(0.047)	(0.052)	(0.043)	(0.048)
2nd wife's first two children are girls	-0.002	0.035	0.022	0.004	-0.008	-0.019
	(0.061)	(0.071)	(0.050)	(0.057)	(0.045)	(0.051)
1st wife GG x 2nd wife GG		-0.152		0.071		0.041
		(0.125)		(0.135)		(0.103)
Observations	1867	1867	1867	1867	859	859
R-squared	0.095	0.096	0.043	0.044	0.064	0.065
Mean	0.049	0.049	-0.086	-0.086	0.532	0.532
Panel B: Second wives						
	index tolerance		**index decision-making**		**anemia**	
	(1)	(2)	(3)	(4)	(5)	(6)
1st wife's first two children are girls	-0.047	-0.023	0.020	0.039	-0.068	-0.077
	(0.061)	(0.070)	(0.045)	(0.052)	(0.045)	(0.052)
2nd wife's first two children are girls	-0.023	0.007	-0.028	-0.005	0.027	0.016
	(0.058)	(0.071)	(0.050)	(0.059)	(0.044)	(0.051)
1st wife GG x 2nd wife GG		-0.120		-0.093		0.040
		(0.129)		(0.112)		(0.102)
Observations	1872	1872	1872	1872	866	866
R-squared	0.107	0.107	0.082	0.082	0.056	0.056
Mean	0.032	0.032	-0.120	-0.120	0.523	0.523

Note: The dependent variables are a score of justifying wife beating, a score of sole decision-making in the household, and the probability to have anemia. GG stands for the first two children of the respondent are girls. I control for women's ethno-linguistic group, religion, age category, the age difference with the partner, the age at first union, the duration of the union, the total number of birth, the DHS survey wave, the husband's education, the region of residence, and the DHS wealth score of the household. For anemia, I also control whether women are pregnant at the time of the survey. I compute robust standard errors. Significance level: * p<0.10, ** p<0.05, *** p<0.01. Sample: Women aged 17-49 years old, married to a polygamous partner and who declare that they are the first wife or the second wife of their partner. Source: DHS Burkina Faso 2003 and 2010.

I have provided evidence in section 2.2 that there did not seem to be a birth stopping rule based on the gender of the children, and Rossi and Rouanet (2015)demonstrated that there was no change in birth spacing behavior of women according to the gender of their children in Burkina Faso. As a result, the rank of the first male birth of a woman can be considered as good as random. In order to make the comparisons as meaningful as possible, I first focus on the subsample of women who had at least one male birth and have four children or less, four children being the mean number of children in the sample. As shown in column (1) of table A-2.17, the FE model yields that for first wives, having a first born son at a higher rank (rank 4 rather than rank 1) increases the index of tolerance of IPV by 56%. It suggests that for first wives the timing of the arrival of the first son has an impact on her ability to condemn a harmful gender norm. For second wives, having a first born son at rank 2 rather than decreases

the index of tolerance of IPV by 46% compared to first wives who welcome a son at rank1, and I can reject the hypothesis that having a boy at rank 2 rather than at rank 1 is equivalent for second wives. I cannot exclude however that the overall impact of having a first son at rank 2 for second wives is not different than zero (pvalue = 0.16).

As shown in column (4) of table A-2.17, for first wives, having two girls then a boy born at rank 3 rather than a boy born at rank 1 increases the index of decision-making by 56% compared to first wives having a son at rank 1. For second wives, having a first born son increases the index of decision making by 47% compared to a first wife with a first born son. And this effect is different than the one of being a second wife with a second born son (pvalue = 0.043), which is overall not statistically different than zero. Table A-2.18 in the appendix shows that the FE estimation shows no significant impact of the rank of the first girl on women's tolerance of IPV, participation in decision-making nor probability to be anemic.

Table A-2.17: Tolerance of IPV and rank of the first boy

	score tolerance IPV			score decision making			anemia		
	FE	LPM	LPM no contrib	FE	LPM	LPM no contrib	FE	LPM	LPM no contrib
Panel: Two cowives of rank 1 and 2									
Ref: rank of the first boy is 1									
boy rank 2	0.213	0.223	0.180	0.114	0.061	-0.035	-0.019	0.027	-0.099
	(0.18)	(0.19)	(0.12)	(0.15)	(0.20)	(0.15)	(0.17)	(0.15)	(0.10)
boy rank 3	0.098	0.221	-0.033	0.318**	-0.004	0.083	-0.143	-0.014	-0.039
	(0.17)	(0.19)	(0.13)	(0.15)	(0.20)	(0.16)	(0.20)	(0.15)	(0.11)
boy rank 4	0.564**	0.247	0.206	0.296	-0.098	0.052	-0.254	0.093	-0.105
	(0.23)	(0.22)	(0.15)	(0.22)	(0.18)	(0.18)	(0.24)	(0.17)	(0.12)
2nd wife	0.019	0.272	0.094	0.267*	-0.033	-0.026	-0.102	0.041	-0.145
	(0.14)	(0.18)	(0.10)	(0.15)	(0.16)	(0.14)	(0.14)	(0.14)	(0.10)
2nd wife x boy rank2	-0.458**	-0.304	-0.181	-0.424*	-0.097	0.077	0.246	-0.066	0.332***
	(0.22)	(0.23)	(0.15)	(0.24)	(0.21)	(0.19)	(0.19)	(0.17)	(0.12)
2nd wife x boy rank3	-0.277	-0.206	-0.110	-0.385**	-0.063	-0.047	0.183	-0.114	0.065
	(0.20)	(0.23)	(0.15)	(0.19)	(0.23)	(0.19)	(0.24)	(0.18)	(0.13)
2nd wife x boy rank4	-0.680***	-0.536**	-0.313*	-0.591	0.539**	-0.003	-0.032	-0.351*	0.155
	(0.25)	(0.26)	(0.17)	(0.40)	(0.27)	(0.22)	(0.23)	(0.20)	(0.14)
Observations	2456	809	1647	2456	809	1647	1091	367	724
Contributing to id	809			809					
R-squared	0.119	0.105	0.109	0.092	0.054	0.067	0.122	0.042	0.056
Mean control	1.002	0.908	1.044	0.564	0.584	0.556	0.587	0.522	0.615
Husbands	2108	646	1462	2108	646	1462			
Contributing to id	646			646					
pval test boy rank2 + 2nd wife x boy rank2 = 0	0.091			0.043					
pval test boy rank3 + 2nd wife x boy rank3 = 0	0.331			0.631					
pval test boy rank4 + 2nd wife x boy rank4 = 0	0.605			0.438					
pval test boy rank2 + 2nd wife + 2nd wife x boy rank2 = 0	0.162			0.779					
pval test boy rank3 + 2nd wife + 2nd wife x boy rank3 = 0	0.432			0.165					
pval test boy rank4 + 2nd wife + 2nd wife x boy rank4 = 0	0.685			0.938					

Note: In columns 1, 4, and 7, I report the results yielded by the husband fixed effect (FE) estimation, and in columns 2, 5, and 8, I report the results yielded by the linear probability model (LPM) on the subsample of respondents who contribute to identifying the husband FE model. In columns 3, 6, and 9, I report the estimates obtained by using a LPM on the subsample of wives who do not participate to the identification of the husband fixed-effect. In the husband FE, I control for women's ethno-linguistic group, religion, age category, the age difference with the partner, the age at first union, the duration of the union, the total number of birth, and the DHS survey wave. For anemia, I also control whether women are pregnant at the time of the survey. I cluster standard errors at the husband level. In the LPM, I also control for the husband's education, the region of residence, and the DHS wealth score of the household. I compute robust standard errors. Significance level: * p<0.10, ** p<0.05, *** p<0.01. Sample: Women aged 17-49 years old, married to a polygamous partner and who declare that they are the first wife or the second wife of their partner. Source: DHS Burkina Faso 2003 and 2010.

Table A-2.18: Tolerance of IPV and rank of the first girl

	score tolerance IPV			score decision making			anemia		
	FE	LPM	LPM no contrib	FE	LPM	LPM no contrib	FE	LPM	LPM no contrib
Panel: Two cowives of rank 1 and 2									
Ref: rank of the first girl is 1									
girl rank 2	-0.117	0.033	-0.088	0.151	-0.158	0.108	-0.033	0.273	0.121
	(0.21)	(0.22)	(0.13)	(0.20)	(0.22)	(0.13)	(0.17)	(0.17)	(0.09)
girl rank 3	0.062	0.318	-0.041	0.305	-0.139	0.195	0.014	0.359**	0.040
	(0.20)	(0.23)	(0.14)	(0.23)	(0.23)	(0.16)	(0.19)	(0.16)	(0.10)
girl rank 4	-0.015	0.198	-0.057	0.152	0.023	0.019	0.077	0.209	0.118
	(0.22)	(0.23)	(0.16)	(0.28)	(0.23)	(0.17)	(0.21)	(0.18)	(0.12)
2nd wife	-0.202	0.183	-0.128	0.159	-0.148	0.107	-0.153	0.115	0.102
	(0.19)	(0.20)	(0.12)	(0.19)	(0.19)	(0.13)	(0.17)	(0.15)	(0.09)
2nd wife x girl rank2	0.062	0.032	0.097	-0.031	0.225	-0.134	0.121	-0.126	-0.184*
	(0.24)	(0.26)	(0.16)	(0.22)	(0.22)	(0.16)	(0.22)	(0.19)	(0.11)
2nd wife x girl rank3	-0.107	-0.226	0.210	-0.136	0.375	-0.180	0.181	-0.209	0.007
	(0.23)	(0.26)	(0.16)	(0.25)	(0.24)	(0.18)	(0.22)	(0.19)	(0.12)
2nd wife x girl rank4	0.058	-0.163	0.157	-0.156	-0.100	-0.081	-0.191	-0.182	-0.215
	(0.27)	(0.26)	(0.18)	(0.25)	(0.24)	(0.18)	(0.23)	(0.21)	(0.14)
Observations	2331	770	1561	2331	770	1561	1062	351	711
Contributing to id	770			770					
R-squared	0.064	0.141	0.108	0.108	0.037	0.037	0.142	0.068	0.056
Mean Control	0.926		1.022	0.586		0.544	0.474		0.508
Husbands	2007	632	1375	2007	632	1375			
Contributing to id	632			632					
pval test	0.739			0.254					
pval test	0.841			0.428					
pval test	0.841			0.428					

Note: The dependent variables are a score of justifying wife beating, a score of sole decision-making in the household, and the probability to have anemia. In columns 1, 4, and 7, I report the results yielded by the husband fixed effect (FE) estimation, and in columns 2, 5, and 8, I report the results yielded by the linear probability model (LPM) on the subsample of respondents who contribute to identifying the husband FE model. In columns 3, 6, and 9, I report the estimates obtained by using a LPM on the subsample of wives who do not participate to the identification of the husband fixed-effect. In the husband FE, I control for women's ethno-linguistic group, religion, age category, the age difference with the partner, the age at first union, whether women are in first union, the duration of the union, the total number of birth, and the DHS survey wave. For anemia, I also control whether women are pregnant at the time of the survey. I cluster standard errors at the husband level. In the LPM, I also control for the husband's education, the region of residence, and the DHS wealth score of the household. I compute robust standard errors. Significance level: * p<0.10, ** p<0.05, *** p<0.01. Sample: Women aged 17-49 years old, married to a polygamous partner and who declare that they are the second wife or the third wife of their partner. Source: DHS Burkina Faso 2003 and 2010.

Table A-2.19: Tolerance of IPV and rank of the first boy monogamous

	score tolerance IPV (1)	**score decision making** (2)	**anemia** (3)
Panel: Monogamous wives			
boy rank 2	0.039	-0.010	-0.065*
	(0.04)	(0.05)	(0.03)
boy rank 3	0.021	0.013	-0.083*
	(0.05)	(0.07)	(0.04)
boy rank 4	0.054	0.042	-0.065
	(0.07)	(0.09)	(0.06)
Observations	5277	5277	2288
R-squared	0.140	0.030	0.030
Mean Control	0.801	0.604	0.504

Note: The dependent variables are a score of justifying wife beating, a score of sole decision-making in the household, and the probability to have anemia. I control for women's ethno-linguistic group, religion, age category, the age difference with the partner, the age at first union, whether women are in first union, the duration of the union, the total number of birth, the DHS survey wave, the husband's education, the region of residence, and the DHS wealth score of the household. For anemia, I also control whether women are pregnant at the time of the survey. I compute robust standard errors. Significance level: * p<0.10, ** p<0.05, *** p<0.01. Sample: Women aged 17-49 years old, married to a monogamous partner. Source: DHS Burkina Faso 2003 and 2010.

Table A-2.20: Tolerance of IPV of wives according to sex composition of the children's husband - polygamous

	tolerance ipv (1)	**score decision-making** (2)	**anemia** (3)
Panel A: Polygamous wives rank 1			
Husband's eldest is a boy	0.005	0.044	0.028
	(0.042)	(0.049)	(0.027)
Observations	2941	2941	1325
R-squared	0.073	0.041	0.034
Mean	1.008	0.635	0.527

	tolerance ipv (1)	**score decision-making** (2)	**anemia** (3)
Panel B: Polygamous wives rank 2			
Husband's eldest is a boy	0.015	-0.070	0.014
	(0.047)	(0.044)	(0.039)
Observations	2159	2159	997
R-squared	0.104	0.050	0.045
Mean	0.999	0.494	0.525

Note: The dependent variables are a score of justifying wife beating, a score of sole decision-making in the household, and the probability to have anemia. I control for women's ethno-linguistic group, religion, age category, the age difference with the partner, the age at first union, the duration of the union, the total number of birth, the husband's education, the region of residence, the DHS wealth score of the household, and the DHS survey wave. I compute robust standard errors. Significance level: * p<0.10, ** p<0.05, *** p<0.01. Sample: Women aged 17-49 years old, and married to a polygamous partner who has exactly two wives. Source: DHS Burkina Faso 2003 and 2010.

Table A-2.21: Tolerance of IPV of wives according to sex composition of children's husband - monogamous

	tolerance ipv (1)	**score decision-making** (2)	**anemia** (3)
Panel B: Monogamous husbands			
Husband's eldest is a boy	0.014	0.014	0.014
	(0.039)	(0.039)	(0.039)
Observations	997	997	997
R-squared	0.045	0.045	0.045
Mean	0.525	0.525	0.525

Note: The dependent variables are a score of justifying wife beating, a score of sole decision-making in the household, and the probability to have anemia. I control for women's ethno-linguistic group, religion, age category, the age difference with the partner, the age at first union, the duration of the union, the total number of birth, the husband's education, the region of residence, the DHS wealth score of the household, and the DHS survey wave. I compute robust standard errors. Significance level: * p<0.10, ** p<0.05, *** p<0.01. Sample: Women aged 17-49 years old, married to a monogamous partner who has exactly two wives. Source: DHS Burkina Faso 2003 and 2010.

CHAPTER 3

USING LIST EXPERIMENTS TO MEASURE INTIMATE PARTNER VIOLENCE (IPV): EVIDENCE FROM RURAL BURKINA FASO

Abstract: I implement a measurement exercise and compare the estimates of the prevalence of intimate partner violence (IPV) obtained with direct questions and with a list experiment (LE), an indirect non-conventional measure for sensitive opinion or behavior. I measure the prevalence of several types of IPV (less severe physical violence, severe physical violence and marital rape) among women living in rural areas of Burkina Faso and find that the direct measures of IPV underestimate by 7 to 9 percentage points the most intense forms of violence. I also find that IPV correlates differently with a series of respondents' characteristics according to the measure of IPV used, including whether the first child of a woman is a girl. To the best of my knowledge, this paper is among the first to study the bias of IPV in a Western African country and whether this bias differs according to the intensity of IPV. This work fuels the debate around the relevance of the current measures of IPV in household surveys that seem to significantly underestimate the prevalence of IPV.

Introduction

Policy makers aiming to improve women's welfare have shown an increasing concern about reducing women's exposure to intimate partner violence (IPV). IPV remains a sensitive topic, however, and its measure is prone to a downward and nonrandom bias (Tourangeau and Yan (2007)). For lack of other reliable sources, especially in developing and low-income countries, many studies rely exclusively on data[1] that is unanimously considered as flawed. Mismeasuring IPV has serious consequences as it impacts the diagnosis of the issue, the design of policies meant to tackle it, and the ability to evaluate how successful policies are at solving the problem.

In this paper, I compare the standard measure of IPV to the measure yielded by a list experiment (LE), in the context of rural Burkina Faso. I explore two main questions: how much of IPV is missed with the standard measure of domestic violence, and who are the women that we miss with the standard measure. The LE was first developed in the 1980's (Miller (1984)). It is designed to conceal the individual binary response to a sensitive question by aggregating it with the answers to several binary innocuous questions. The answer to the sensitive item is thus hidden and provides respondents a certain degree of anonimity or at least plausible deniability when asked whether they were the victim of IPV. By randomizing lists containing a sensitive item and lists containing non-sensitive items, researchers can then identify prevalence of IPV. The LE is expected to decrease bias but at the cost of higher variance. I first assess whether the prevalence of IPV obtained with the LE differs from the one yielded by direct questioning. I also go a step further by checking whether the difference in prevalence of IPV with the LE and direct questions varies according to the nature of the violence measured, investigating less severe physical violence, physical violence and marital rape. Finally, I study whether some respondents' characteristics correlate differently with IPV according to the method used to measure it.

I find that women tend to under-declare their experience of severe physical violence and marital rape by 7 to 9pp respectively when asked directly compared to what is yielded with the LE. I find no statistically significant difference when it comes to less severe physical violence. Unlike the direct measure of IPV that suggests that fewer women experience more intense forms of IPV, the LE measure yields that women seem equally exposed to the three types of IPV. The results also suggest that women's age, sex of the eldest child and exposure to a mass media

[1] Many studies rely on the Demographic and Health Surveys (DHS), which a precious resource in studying IPV in the absence of other data on the topic at least for the African continent.

campaign meant to encourage modern contraception uptake correlate differently with experience of IPV according to the method used. I notably find that mothers of first born daughter are 22pp more likely to report marital rape with the list, whereas I detect no significant effect with the direct measure. I also find descriptive evidence that being exposed to the media campaign is positively correlated with marital rape measured with the LE, whereas no effect is detected with the direct measure.

This paper contributes to a growing literature on measurement of sensitive behaviors by comparing LE and direct measures in contexts where there are little to no alternative to the DHS measures of IPV (Cullen (2020) in Rwanda and Nigeria, Treibich and Lépine (2016) on condom use in Senegal, Peterman *et al.* (2018), Bulte and Lensink (2019) Agüero and Frisancho (2017) in Peru and Chuang *et al.* (2021) in Côte d'Ivoire). Unlike much of the literature, however, I rely on data that are representative of women living in rural areas in all of Burkina Faso rather than women living in one specific region of a country. While Agüero and Frisancho (2017) find no systematic difference in IPV declaration under the LE and direct measure in Peru, I find, like Cullen (2020) in Rwanda, that women under-declare IPV when asked directly. However, while she finds that women in Nigeria declare equally high experience of marital rape when asked directly or with the LE, I find that women are more likely to declare marital rape with the LE. The difference in our results may stem from contextual differences: Cullen (2020) underlines that qualitative work she did suggested that there was less stigma associated with sexual IPV than physical IPV in the context she studies, whereas qualitative work I led in Burkina Faso suggest that, even though women recognize that having intercourse is a husband's prerogative, exerting violence to obtain it remains frowned upon.

While much of the literature underlines association between women's characteristics and differences in reporting of IPV (like education or bargaining power, see Cullen (2020) or Agüero and Frisancho (2017)), I am able to exploit exogenous variations in the sex of the first child to investigate the potential existence of a preference for boys in rural Burkina Faso, and to study more comprehensively the impact of an "edutainment" program on IPV using two different measures of violence, albeit in a descriptive manner. I also provide a methodological contribution as it is the first empirical paper to include in the design of the LE safeguards suggested by econometricians that perfected the technique (Droitcour *et al.* (2004), Glynn (2013), Imai (2011), Blair and Imai (2012), Blair *et al.* (2018), Aronow *et al.* (2015), Chou *et al.* (2017)) to assess the consistency of the LE.

Finally, this paper may contribute to understanding the mixed results of interventions aiming at changing gender-norms in order to decrease gender-based violence in general and IPV in particular. As formalized by Bénabou and Tirole (2006) and Benabou and Tirole (2011), norm-based interventions may spur social change by notably harnessing the utility agents derived from the social image. However, according to models, the potential for impact of norm-based interventions also depends on the distribution of social preferences or the frequency of the prosocial behavior in the community. Norm-based interventions may mistakenly rely on the assumption that abstaining from abusing one's spouse falls into the category of "respectable" behavior that only "the worst people" do not engage in, whereas the prevalence of IPV yielded by the LE suggests that, in the context studied here, abstaining from abusing one's spouse might actually be closer to a "modal" behavior, for which Bénabou and Tirole (2006) and Benabou and Tirole (2011) suggest that norm-based interventions may not be as effective as other policies relying on higher incentives such as cash-transfers for instance.

The paper is organized as follows. Section 3.1 provides an econometric justification for using list experiments as it is meant to reduce the bias in the estimation of the impact of any treatment effect on the experience of IPV induced by nonclassical measurement errors. Section 3.2 describes the context of the implementation of the LE and the experimental design. Section 3.3 details the empirical strategy. Sections 3.4 and 3.5 respectively provide a series of validity checks and a conceptual framework to apprehend the results presented in Section 3.6. Section 3.7 provides a discussion of the results and of the relevance of the methodological tools available to exploit LE data. Section 3.8 concludes.

3.1 The Case for Implementing List Experiments: a Simple Econometric Model

The presence of nonclassical error in the measurement of an outcome has impacts on our ability to assess the impact of a treatment on the outcome. I present a simple model to illustrate how nonclassical errors, that are likely to arise when surveying women about IPV, impact this process.

I am interested in the impact of X on Y^*, so I run the following linear regression: $Y^* = \beta * X + u$. I assume that X is exogenous to Y^* so that $cov(X, u) = 0$ and $E(u|X) = 0$. I cannot perfectly observe the experience of IPV of women Y^*, but I observe the report of the experience of IPV:

Y, with a level of error v, so that:

$$Y = Y^* + v \tag{3.1}$$

I assume that the measurement error v consists in two components: a random component, v', and a component that is correlated with Y^* and with the variable X is a linear way, so that:

$$v = \rho * Y^* + \delta * X + v' \tag{3.2}$$

In practice, I run the following regression: $Y = \beta_b * X + u$. By definition:

$$\beta_b = \frac{cov(Y, X)}{V(X)}$$

Substituting Y for Y^* and v:

$$\beta_b = \frac{cov(\beta * X + u + \rho * Y^* + \delta * X + v'), X)}{V(X)}$$

$\iff \beta_b = \beta + \rho * \frac{cov(Y^*,X)}{V(X)} + \delta * \frac{cov(X,X)}{V(X)}$
$\iff \beta_b - \beta = \rho * \beta + \delta * \frac{cov(X,X)}{V(X)}$
$\iff \beta_b - \beta = \rho * \beta + \delta$

I assume that the nonrandom part of the error in measurement increases with experience of IPV, i.e. that $\rho \geq 0$.

If

$$\beta_b - \beta < 0$$

$\implies$

$$\rho * \beta + \delta < 0$$

$\Longrightarrow$

$$\rho * \beta < -\delta$$

if $\rho = 0$, then

$$\delta < 0$$

, i.e. the measurement error induced by the treatment variable is negative.
if $\rho > 0$, then

$$\beta < \frac{-\delta}{\rho}$$

, i.e. the true effect of the treatment on IPV is inferior to minus the ratio of the measurement error induced by the treatment variabe, and the measurement error induced by the true experience of IPV. Conversely, if $\beta_b - \beta < 0$, then $\beta > \frac{-\delta}{\rho}$.

The model has two main takeaways: first, it shows that even in the absence of a nonrandom correlation between the true experience of IPV and the error term (i.e. the case where $\rho = 0$), the impact of X on Y remains biased by δ, whose sign is not obvious to predict. Second, if $\rho > 0$, then the direction of the bias of the impact of X on Y is determined by a ratio for which there is no obvious prior as to the sign and magnitude of its numerator and denominator. Consequently, any survey instrument that may help alleviate the nonclassical measurement error that is bound to arise when surveying women on IPV is a welcome effort to refine our understanding of the impact of public policies on women's experience of domestic violence. The list experiment is one possible tool as it is designed to reduce nonclassical measurement error by reducing the bias in the reporting of violence among women surveyed.

Section C-3.1. in the Appendix provides a formula of the bias for when the independent variable X has heterogeneous effects.

3.2 Context, Design of the Experiment, and Data

3.2.1 Context of the implementation

In the context of sub-Saharan Africa and West Africa more particularly, there are reasons to believe that direct measures of IPV are likely to be affected by a non-negligible bias. First, the question can be perceived as intrusive as it touches upon very intimate aspects of a woman's life. Second, in the tight-knit rural societies of West African countries, women who are victim of IPV may fear that their answer may come to be known by their partner, which would expose them to retaliation. Third, though IPV is tolerated to a certain extent,[2] anthropological work suggests that, out of loyalty for their husband, no matter what happens behind closed doors, women are expected to salvage their husband's image and social reputation (Falen (2011)). Additionally, women may not want to appear as victims or as having had to be disciplined by their husband, which could suggest they allegedly did something wrong. Therefore, the social pressure exerted on women to keep their experience of IPV quiet is expected to be strong.

The list experiment was added to the follow-up survey of a study on the impact of a mass media radio campaign on family planning in rural Burkina Faso.[3] The Program was a two and a half years radio campaign implemented between June 2016 and December 2018. 10 one-minute radio spots per day and 2 regular mocked phone-in radio shows per week were broadcasted in vernacular languages. The themes adressed in the radio shows went beyond the question of contraceptive use: they ranged from the drawbacks of early marriage to fostering discussion between spouses around contraceptive use. The intervention is a clustered randomized controlled trial (RCT) with 16 radio stations. In addition, an individual treatment was implemented as 1500 women were randomly selected to receive a radio. There are 2 waves of survey data from 250 villages: one collected in 2016 - without a list experiment but with baseline data on experience of IPV collected through direct questioning - and a follow-up in 2018. Because the family planning intervention may have an impact on IPV, I stratified the list experiment at the village and individual radio treatment level.

Going through the questionnaire of the follow-up survey lasted an hour to an hour and a half, including the list experiment. Given the sensitive nature of the whole questionnaire, the survey

[2] According to the 2010 DHS for Burkina Faso, 40% of women found it justifiable for a husband to beat his wife if she goes out without telling him, neglects the children or argues with him.

[3] "The Media or the Message? Experimental Evidence on Mass Media and Modern Contraception Uptake in Burkina Faso" with Rachel Glennerster and Joanna Murray. January 2021

was administered by female enumerators only. Respondents were told from the beginning that the data was collected by the NGO Innovation for Poverty Action (IPA) to inform health and contraception-related behaviours.

3.2.2 Design

The principle of the list experiment consists in presenting one group of individuals (group A) a list of three non-sensitive items, for instance, and to ask them with how many items they agree in total. Then, another group (group B) is presented the same list of three innocuous items + a sensitive item, which is the sensitive opinion or behavior the researcher wants to measure. By virtue of the randomization, the difference in the mean number of items agreed upon in group B minus in group A should reveal the true prevalence of the sensitive opinion or behavior.

At the beginning of the section on IPV, respondents were read a text introducing the game: *I'm going to read 3/4 sentences describing situations that may happen to women. I'm going to read those sentences one after the other. When I'm done, what I'm interested in is to know HOW MANY sentences apply to your situation. When I read a sentence, do not tell me "yes" or "no". I just want to know how many sentences in total apply to your situation.*

To help you remember your answer, I am going to give you 3/4 marbles. Put those marbles in your right hand, close your hands and put them both in your back like this ***[The enumerator shows the respondent how to put her hands in her back]****. When I read a sentence, for instance "I like to wear red", if you disagree with the sentence, you do nothing* ***[the enumerator shows her hands in her back and does nothing]****. If you agree with the sentence, pass a marble from your right hand to your left hand, like this [the enumerator shows how to pass the marble]. In the end, take your hands off your back and show me the marbles you have in your left hand only. To give you more confidentiality, I am going to turn a bit during the game* ***[the enumerator turns her body at a 90 degrees angle.]***

Marbles were provided to the respondent to help her keep track of her answers. After explaining the game to the respondent, the enumerator would then use a "dummy" list to make sure the respondents understood the protocol. The respondent was told that the dummy list was meant to test whether she understood the protocol. If she understood the protocol, the enumerator would move on to the actual three lists of interest: the list containing the less severe physical violence item, the list containing the severe physical violence item, and the one with

the item about conjugal rape.

List1 + less severe physical violence: *I am going to read 3/4 sentences, one after the other. If you agree with the sentence I read, pass a marble from your right hand to your left hand. Otherwise, do nothing, just like we trained to do it.*

- I always eat alone. *If this sentence applies to your situation, pass a marble from your right hand to your left hand, if not, do nothing.*
- Last week, my husband attended a funeral. *If this sentence applies to your situation, pass a marble from your right hand to your left hand, if not, do nothing.*
- **My husband/partner has already slapped me or thrown at me an object that could have hurt me.** *If this sentence applies to your situation, pass a marble from your right hand to your left hand, if not, do nothing*
- I dedicate more time than my husband to domestic chores. *If this sentence applies to your situation, pass a marble from your right hand to your left hand, if not, do nothing.*

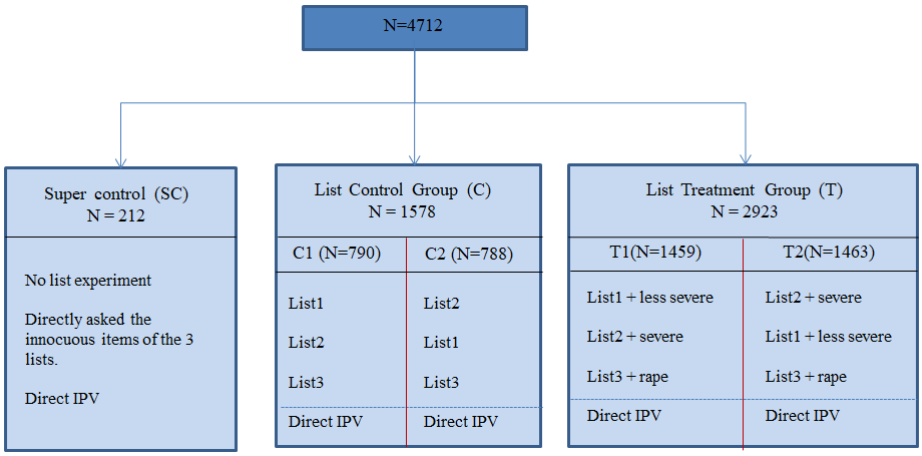
List2 + severe physical violence:

- Every day, my husband sweeps the courtyard. *If this sentence applies to your situation, pass a marble from your right hand to your left hand, if not, do nothing.*
- I regularly go to the market. *If this sentence applies to your situation, pass a marble from your right hand to your left hand, if not, do nothing.*
- **My husband/partner has already punched me, kicked me or dragged me to the ground.** *If this sentence applies to your situation, pass a marble from your right hand to your left hand, if not, do nothing.*
- Last week, my partner/husband got so sick he couldn't work. *If this sentence applies to your situation, pass a marble from your right hand to your left hand, if not, do nothing.*

List3 + conjugal rape:

- Last week, I paid someone to do my husband's laundry. *If this sentence applies to your situation, pass a marble from your right hand to your left hand, if not, do nothing*
- My husband/partner often takes his dinner at home. *If this sentence applies to your situation, pass a marble from your right hand to your left hand, if not, do nothing.*

Figure 3.1: Design of the experiment



Sample: Women married or living with with partner at the time of the survey.

- **My husband/partner has already forced me to have sex with him when I didn't want to.** *If this sentence applies to your situation, pass a marble from your right hand to your left hand, if not, do nothing.*

- Last week, I bought myself a new wax or a new outfit. *If this sentence applies to your situation, pass a marble from your right hand to your left hand, if not, do nothing.*

Respondents were randomly allocated to five groups; in a "super control" group (SC), women received no list experiment but were instead directly asked whether the innocuous items of the list applied to their situation and were then asked directly about IPV. In the treatment group (T), women were split between two groups: T1 and T2. T1 received the 3 lists in the following order; first the list1 + the less severe physical violence item, list2 + the severe physical violence item and eventually list3 + the conjugal rape item. T2 was first read list2 + the severe physical item, then list1 + the less severe physical item and eventually list3 + the conjugal rape item. The control group (C) was split between C1 and C2 and were read list1, list2 and list3 in an order mirroring (T1) and (T2) respectively. Figure 3.1 summarizes the different groups created and the treatment received. All respondents were asked about their experience of IPV after the list. Table A-3.4 in Appendix shows the randomization was successful.

3.2.3 Ensuring the quality of the IPV data

The ability to capture a difference in the prevalence of IPV according to the measure used hinges on the quality of each measure.

Direct measure of IPV

Following standard guidelines to protect women who answer questions about IPV, enumerators were instructed to isolate the respondent and to make sure that no one could hear the questions asked when inquiring about experience of IPV. To build trust more easily, and given the sensitive nature of the questions, enumerators were all women, and they were specifically selected, grouped and dispatched to areas of similar ethno-linguistic groups as their own, or if they did not belong to the same ethno-linguisitic group as respondents, they spoke fluently the respondent's language. The survey module on domestic violence came at the end of a follow-up survey on contraception use and sexual activity. It means that, when the enumerators reached the IPV section of the survey, they have already covered intimate topics with the respondent and build a rapport with them, which contributes to help women open up about their experience.

List experiment

For the list experiment, the same protocol applied but additional precautions were taken in order to ensure the quality of the list experiment protocol. Because 80% of women declared to have no formal education in the baseline survey, marbles were used to help them keep track of the number of sentences they agree with. The use of marbles was also implemented to avoid the phenomenon of "satisfecing". "Satisfecing" refers to the propensity of the respondent to give what she or he believes to be an acceptable number of answers when asked with how many sentences they agree with, rather than providing the actual answer. Andrew (2014) describes this issue experienced on the field and suggests ways to remedy it that are similar in spirit as the use of marbles. In addition, the protocol states that the enumerator is the one counting and reporting the number of marbles in the respondent's left hand. Because list experiment is more cognitively taxing than usual survey instruments like direct questions, respondents were presented with four lists, including a dummy list to test the understanding of the respondent before moving to the three lists of interest.

Finally, the training of the enumerators is crucial to ensure the quality of the data collected with the list experiment. Prior to the survey implementation, all enumerators participated in an intensive six-day training. Two days were dedicated to the IPV module: a one-day classroom session and an extensive practice session. I personally took part in the training sessions and prepared materials for the training. For the classroom sessions, I used three different types of

support material to ease enumerators' understanding of the list experiment protocol. I first used the classic enumerator's training manual put together by IPA's team. Second, I created a photo-story that displayed how to conduct the list experiment, detailed its script, and insisted on the dos and don'ts of the protocol. Each enumerator was handed out a copy of the photo-story to study at home. Finally, enumerators were shown three times a video of two women[4] performing the list experiment. After showing the video, two of the enumerators who took part in the pilot study of the experiment (and thus knew well how to conduct it) were invited to show the other trainees how to conduct the list experiment. Then, trainees were paired and asked to practice the list experiment. During the practice session, the respondents were taking turns to interview a consultant standing as a respondent who assessed their performance. Only a subsamble of women got to practice the list experiment protocol. The others were gathered by teams (correponding to an ethno-linguistic group). For each group, I stood as a respondent and one the enumerator practiced the list experiment with me in French. Once the delivery of the script was satisfactory, two other enumerators of the same team were invited to practice the LE protocol in the language in which they were to conduct the interviews, in front of the other members of their teams. The team was then invited to make comments on what was well-done by their teammates and what needed to be improved. Then another pair of enumerators would practice in front of the others and receive comments for improvements until all members of the group practiced as the enumerator. A week after the training, team leaders received an additional day of training, during which they watched again the video of the LE protocol and were reminded of the key steps that must be respected in the protocol.

3.2.4 Sample

Table 3.1 reports descriptive statistics for the analysis sample. It shows that women are on average nearly 34 years old,[5], 59% of them belong to the main ethno-linguistic group, the Mossi and 65% of them are muslim. They started their marital life at nearly 18 on average and gave birth for the first time at 19 years old. 53% of them were married to a polygamous partner at the time of the follow-up survey. Half the women in the sample believe it is acceptable for a husband to beat his wife if she goes out without telling him or neglects the children. 60% of women in the sample find wife beating acceptable if a woman refuses sex to her spouse.

[4] I kindly thank Juliette Crespin-Boucaud, PhD student at PSE at the time, for her appearance as the respondent in the video

[5] The age range for the endline survey is 18 to 52.

Table 3.1: Sample descriptives

VARIABLES	mean	sd	N
age	33.95	8.81	4,712.00
Husband is polygamous	0.53	0.50	4,712.00
respondent earns money	0.48	0.50	4,712.00
age at first child	19.21	2.44	4,313.00
wife beating tolerable for going out without telling husband	0.50	2.34	4,712.00
wife beating tolerable for neglecting the children	0.53	1.83	4,712.00
wife beating tolerable for refusing sex	0.61	4.52	4,712.00
wife beating tolerable for burning the food	0.30	2.62	4,712.00
age at first union	17.75	2.45	4,301.00
respondent has some formal education	0.18	0.38	4,712.00
husband has some formal education	0.25	0.43	4,574.00
muslim	0.65	0.48	4,712.00
mossi	0.59	0.49	4,712.00
eldest child is a girl	0.51	0.50	4,591.00
baseline physical ipv	0.15	0.36	4,712.00
baseline marital rape	0.06	0.23	4,712.00

Note: Table reports descriptive statistics for the super control group, the control groups and the treatment groups.

3.3 Empirical Strategy

For now, I estimate the prevalence of IPV using a simple difference-in-means estimator. The following regression provides an estimate of the prevalence of IPV measured with the LE.

$$Y_i = \alpha + \beta_s + \theta * LIST_i + \delta * X_i + \varepsilon \tag{3.3}$$

with Y_i the mean number of statements with which the respondent agrees when read a givel list, β_s a strata fixed-effect and $LIST_i$ a dummy variable indicating whether the respondent was read the series of list that included the sensitive items. θ is our coefficient of interest. X_i is a series of individual controls including respondent's age category, whether she has some formal education, belonging to the main ethno-linguistic group (i.e. being a mossi), and belonging to the main religious group (i.e. being a muslim). Standard errors are clustered at the strata level.

The estimates of IPV from direct questioning are obtained by regressing respondent's binary answer to whether they were the victim of less severe or severe physical violence or marital rape on a constant. Estimates of prevalence of IPV yielded by the LE and by the direct questioning are then compared using a seemingly unrelated regression (with the suest STATA command).

I will further investigate the characteristics of women who declare IPV with the LE using the

following regression:

$$Y_i = \alpha + \beta_s + \theta * LIST_i + \sum_{j=1}^{J} \gamma_j * S_{ji} + \sum_{j=1}^{J} \rho_j * LIST_i * S_{ji} + \varepsilon \tag{3.4}$$

where S_{ji} is characteristic j of respondent i (and J the total number of characteristics), such having some formal education, belonging to a given age category, belonging to the main ethno-linguistic group (i.e. being a mossi), and belonging to the main religious group (i.e. being a muslim).

In order to assess whether there is a significant difference in the impact of each correlates according to the measure of IPV used, I use the following regression to measure the impact of the different correlates of interest on the direct measure of IPV before using a seemingly unrelated regression to compare δ and ρ_j for each variable of interest:

$$Y_i = \alpha + \beta_s + \delta * X_i + \varepsilon \tag{3.5}$$

3.4 Validity Checks

3.4.1 Design, floor and ceiling effects

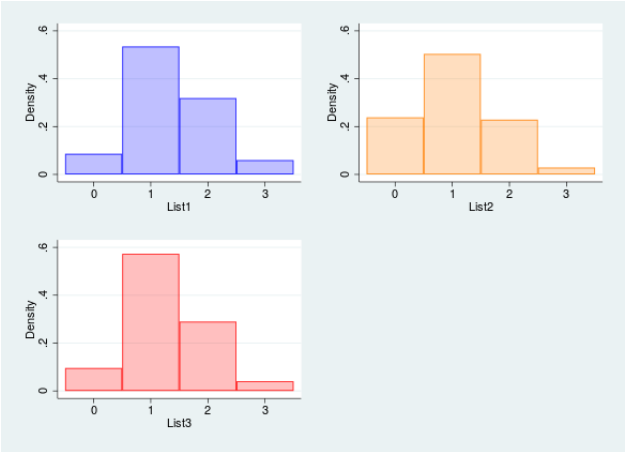
The estimates yielded by the LE are valid under the assumptions of no design effects and no liars. As defined in Blair and Imai (2012), design effects occur when the inclusion of the sensitive item to the baseline list affects the answer of the respondents to the baseline items. Two tests are available to assess the robustness of the data to this effect. The first one is the most usual in the literature and was suggested by Glynn (2013). It consists in checking whether the difference in the share of respondents who answered "yes" to at least j items (with $j = 1, 2, 3, 4$) in the treatment group and in the control group is positive for all j. As table A-3.1, table A-3.2 and table A-3.3 in Appendix, the proportion of interest are always positive, which suggests there is no design effect as described byBlair and Imai (2012).

The second test was developed by Blair and Imai (2012). The principle of the test is the following: under the two assumptions of no design effect and no liars, the joint distribution of the number of non-sensitive item and the true preference for the sensitive item is identified.The test estimates the proportion of respondent that fall in each possible cell representing a pair of control item count (from 0 to 3) and the truthful preference for the sensitive item (0 or 1). When those counts are estimated to be negative, it indicates that respondents' answers to the control

item counts differ across control and treatment groups thus violating the no design effect assumption. All three lists pass the test computed with the R package proposed by Blair and Imai (2012) as table A-3.6, table A-3.7 and table A-3.8 in Appendix show it.

Following Glynn (2013), I check whether many respondents answered "yes" to all the baseline items (ceiling effects) or to none of them with high frequency (floor effects). Floor or ceiling effects may undermine the quality of the estimates yielded by the list because respondents may feel that their answer to the sensitive item is not properly concealed by the list design. To avoid them, Glynn (2013) encourages to built baseline lists with negatively correlated innocuous items. I opted for the second best option put forward by the literature which was to test the frequency of items in a pilot study and build the list with one item that is rather likely to have happened to respondents, another one which is unlikely to have happened and one whose occurence is uncertain. Figure 3.2 shows that list1 and list3 exhibit no floor nor ceiling effect. Even though list2 exhibits a floor effect (nearly 23% of women had zero marbles in their left hand after being read that list), let's keep in mind that admitting the sensitive answer is tantamount to giving a positive answer. As a result, despite the floor effect, women's answer in the treatment group is still somehow protected, even though it is to a lesser extent than with the other two lists. Additionnaly, as Zigerell (2011) mentions it, the bias it induces is predictable and bias the LE estimates downwards.

Figure 3.2: Count of answers in the Control Group per list



Note: sample: control groups only. N = 1578.

Finally, Chuang *et al.* (2021) argue that list experiments build with innocuous baseline statements only makes the sensitive statement of interest very obvious and salient. As a result, such list experiments may be jeopardized by another type of design effect than the one men-

tionned by Blair and Imai (2012). The design effect they underline may happen if respondents are cued and believe that researchers can somehow retrieve their answer to the sensitive item even though they cannot. The point they make speaks to the respondent's trust in the ability of the protocol to conceal answers to the sensitive statement. Though I won't argue against the fact the items about exprience of IPV are indeed salient among the innocuous baseline items chosen, the protocol designed here still gives women plausible deniability about their answer to the sensitive item. Figure 3.3 shows the distribution of innocuous items in the super control group. The patterns of answers in each list suggest that chosen innocuous items vary enough to provide plausible deniability to women about their experience of IPV.

Figure 3.3: Distribution of innocuous items

Note: Sample: super control group only. N = 212.

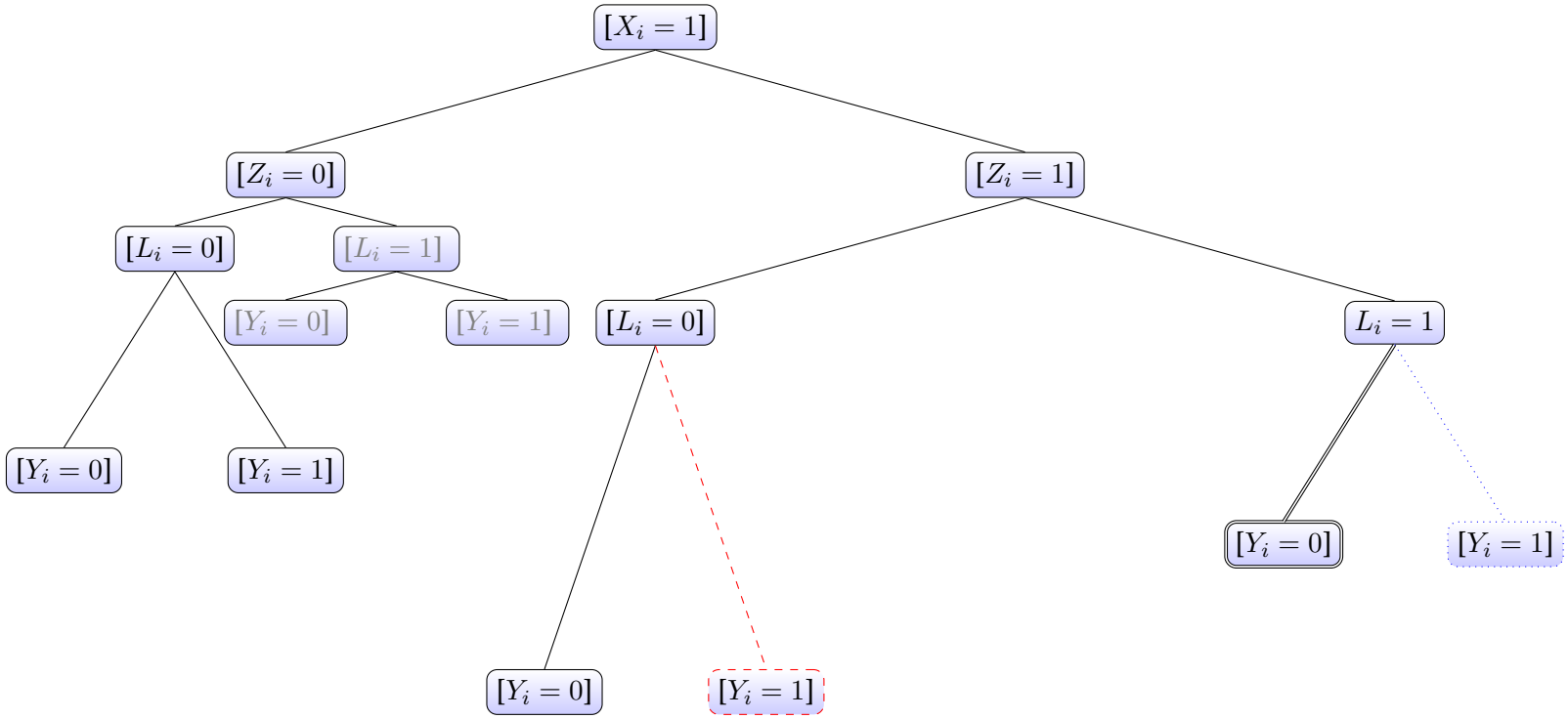
I also provide additional consistency checks that are less usual in the literature. In particular, I provide evidence that the mean number of innocuous items agreed upon are not statistically significantly different in the super control group and in the group that received the baseline lists (figure A-3.1 in Appendix).

3.5 Framework

In order to clarify the possible pattern of answers of respondents I suggest a simple framework drawn from Chou (2018). Some women have experienced IPV, others have not ($X_i \in \{0, 1\}$). I randomly expose women to a treatment, which consists in being exposed to sensitive items about IPV with a list experiment ($Z_i \in \{0, 1\}$). Then, women have two opportunities to reveal

whether they experienced IPV. The first time, they are asked to reveal whether they experienced IPV with the list ($L_i \in \{0, 1\}$). Women comply with the list protocol if they experienced IPV and reveal it with the list or if they did not experience IPV and say they were not the victim of it through the list. The second time, all women are exposed to the sensitive item through the direct question ($Y_i \in \{0, 1\}$). At this stage, women who comply with the protocol are those who experienced IPV and reveal it when asked directly or women who did not experience it and say they were not the victim of it. In this framework, summarized in table 3.2, I make the assumption that women who experienced IPV may choose not to reveal it but that women who were not the victim of IPV will not say that they were.

Figure 3.4: Framework - Women who experienced IPV



Note:

$X_i = \begin{cases} 1 & \text{if individual i has experienced IPV} \\ 0 & \text{otherwise} \end{cases}$

$Z_i = \begin{cases} 1 & \text{if individual i randomly assigned to the baseline list + the sensitive item} \\ 0 & \text{if individual i randomly assigned to the baseline list} \end{cases}$

$L_i = \begin{cases} 1 & \text{if individual i reveals experience of IPV with the list} \\ 0 & \text{if not} \end{cases}$

$Y_i = \begin{cases} 1 & \text{if individual reveals she has experienced IPV with the direct question.} \\ 0 & \text{if not} \end{cases}$

Empirically, L_i can only be known in expectation. Women who are not exposed to the baseline list + the sensitive item cannot declare with the list they were the victim of IPV. This is why, events that are considered as null are in gray.

Table 3.2: Categories of respondents who experienced IPV according to their pattern of answers

$X_i = 1, Z_i = 1$	
List-reveal - direct-reveal	$L_i = 1, Y_i = 1$
List-reveal - direct-conceal	$L_i = 1, Y_i = 0$
List-conceal - direct-conceal	$L_i = 0, Y_i = 0$
List-conceal - direct-reveal	$L_i = 0, Y_i = 1$
$X_i = 1, Z_i = 0$	
Direct-reveal	$Y_i = 1$
Direct-conceal	$Y_i = 0$

3.6 Results

3.6.1 Differences in Prevalence of IPV

As shown in table 3.3, 21% of women report that they have already been slapped or thrown an object at by their partner with the list experiment, 20.8% report they were punched, kicked or dragged to the ground by their partner, and 21.2% report that they were raped by their partner.[6]

Table 3.3: Estimated prevalence of IPV with LE

VARIABLES	(1) less_severe	(2) severe	(3) rape
Treated with list	0.209***	0.207***	0.211***
	(0.0418)	(0.0396)	(0.0376)
Observations	4,500	4,500	4,500
R-squared	0.241	0.227	0.181
Controls	Yes	Yes	Yes
Mean	1.478	1.175	1.414

Note: Regression yielding the mean number of marbles in women's left hand in the treated group minus the mean number of marbles in women's left hand in the control group. Controls: strata fixed-effect, age, some formal education, muslim, and mossi. Significance level: $***p < 0.01, **p < 0.05, *p < 0.1$.

The list experiment and the direct measure yield statistically different prevalence of severe physical violence and marital rape, as shown in table 3.4: the prevalence of severe physical violence with the LE is 6.94 percentage points higher than with the direct measure, though it is only marginally significant.[7] The gap in the prevalence of marital rape is 9.2 percentage points and the difference is significant at the 1% level. The direct measure of less severe physical vio-

[6] It is worth noting that the estimates yielded by the panel comparing T1 and C1 and the one comparing T2 and C2 are not statistically different from one another (see table A-3.9 in Appendix). It suggests that, in this study, women are not sensitive to the order in which the sensitive items embedded in the list are presented to them.

[7] at the 14% level.

lence does not seem to be biased compared to the list experiment estimates.

Measuring IPV with the list experiment suggests that women are equally victim of less severe physical violence, physical violence and marital rape during their lifetime. In contrast, using the direct measures of IPV conveys the impression that the more severe the type of violence, the less prevalent.

Table 3.4: Estimated prevalence with LE and direct questions among control groups only

	List Experiment (T - C)	Direct estimated on (C)	Diff	p-value	N
less severe	20.86	17	3.86	0.3284	4500
se	0.039	0.019			
severe	20.68	14.59	6.09	0.1304	4500
se	0.037	0.019			
rape	21.13	11.77	9.36	0.0081	4500
se	0.035	0.018			

Note: First column: output of a regression yielding the mean number of marbles in women's left hand in the treated group to the mean number of marbles in women's left hand in the control group. Controls: strata fixed-effect, age, some formal education, muslim, and mossi. Second column: output of regressing the direct question outcome on a constant.

As women are asked twice about their experience of IPV (once with the list and once directly), I estimate the exact share of women who took advantage of the plausible deniability provided by the list experiment to declare they experienced IPV but switch their answer to ”no” when asked directly. These women can be found along the double black line on figure 3.4 in section 3.5. Table 3.5 confirms that the more intense the form of violence, the more likely women are to conceal their experience of IPV when asked directly. I note that the number of women who switched their answer from “yes” with the LE to “no” with the direct measure is superior to the number of women that would fill the gap between the estimates obtained with the list experiment and with the direct measure shown in table 3.4. This apparent discrepancy can be explained by the presence of women belonging to the category “list-conceal - direct reveal” as outlined in table 3.2. This pattern of behavior is explored in detail in another section.

3.6.2 Policy implications of getting IPV prevalence wrong

Getting the level of IPV prevalence has evident consequences on the assessment of how widespread the phenomenon is, but it may also lead to use tools that may not be the most adequate to tackle it. Bénabou and Tirole (2006) and Benabou and Tirole (2011), develop theoretical models to as-

Table 3.5: Estimated prevalence with LE among respondents who said "no" to the direct question

VARIABLES	(1) less severe	(2) severe	(3) rape
Treated with list	0.0892*	0.145***	0.148***
	(0.0469)	(0.0460)	(0.0397)
Observations	3,585	3,835	3,719
R-squared	0.253	0.250	0.194
Controls	Yes	Yes	Yes

Note: Sample: Women who said they did not experience IPV when asked directly. Dependent variable: Number of items agreed upon in the list experiment. Controls: strata fixed-effect, age, some formal education, muslim, mossi. Standard errors clustered at the strata level in parenthesis. Significance level: $***p < 0.01, **p < 0.05, *p < 0.1$.

sess the relevance of leveraging social stigma or honor to foster the adoption of a prosocial behavior. They argue that norm-based interventions may spur social change by harnessing agent's intrinsic and reputational motivation. In Benabou and Tirole (2011), agents engage in a prosocial action based notably on the "honor" attached to participating, and the "stigma" attached to abstaining from the action. Social pressure is shaped by honor and stigma, which are determined by the distribution of social preferences of what is done/acceptable in the community. The equilibrium behavior can either be (i) "respectable" or "normal", i.e. those not abstaining from beating their wives are the worst people in the community of reference, (ii) "admirable" or "heroic", and performed by the best of the community only, or (iii) "modal", which means that both participating and abstaining are common behaviors. In the modal case, the social pressure to engage in the prosocial behavior is the weakest and trying to leverage agents' social image may not be as effective as monetary incentives. In this work, I show that the direct measure of IPV may mistakenly lead to believe that abstaining from abusing one's spouse falls into the category of "respectable" behavior whereas the prevalence of IPV yielded by the LE suggests that, in the context studied here, abstaining from abusing one's spouse might actually be closer to a "modal" behavior, for which norm-based interventions may not be as effective as cash-transfers for instance.

3.6.3 Heterogeneity in declaration with the list

I now study whether some women are more likely than others to use the list experiment to declare they were the victim of violence. I use the specification detailed in equation 3.4 and interact the exposure to the sensitive item with the list experiment with variables that may have an impact on women's exposure to IPV, or their tendency to reveal IPV, based on associations previously identified by the literature. I explore the relation with the following variables: whether the woman lives in a cluster included in the radio campaign, whether her fist child is a girl,[8] the ethno-linguistic group of women, the religious group, the education level, the age category, and whether women are married to a polygamous partner.

As shown in table 3.6, I find that the radio treatment has no differential impact on women's declaration of less severe and severe physical violence. I note, however, that the coefficient on the variable "campaign clusters" is positive and significant (13.5pp), suggesting that among women who were not treated with the list, those residing in campaign clusters declared on average that they agreed with more innocuous items than women residing in noncampaign clusters. As a result, within campaign clusters, when I difference out the mean number of items agreed upon by respondents in the treatment group and by respondents in the control group, the prevalence of IPV is mechanically downplayed compared to the prevalence of IPV in noncampaign clusters. It may explain why I do not capture a statistically significant in prevalence of physical violence between campaign and noncampaign clusters. By constrast, the list experiment used to measure marital rape exhibits no such flaw, as residing in a campaign cluster has no impact on the number of innocuous items agreed upon (column 13 of table 3.6). I find that women who reside in campaign clusters are 9.36pp more likely to declare they experienced marital rape than women residing in noncampaign clusters.

I also find evidence that the probability to have experienced marital rape among women whose first child is a daughter is 21.8pp higher than among women whose first child is a boy (column 16 of table 3.6). However, the result is not robust to correcting for MHT. In addition, the prevalence of marital rape among women whose first born child is positive yet insignificant, suggesting that, for this group, there may be little interest in the plausible deniability provided

[8] In a context where selective sex abortion is unlikely, the sex of the first child is as good as random. The literature in anthropology has demonstrated that in patrilineal societies in West Africa, the sex of the first born has an influence on women's status in their household as mother of boys contribute to the lineage of their husband (Lesthaeghe (1989)). Additionally, there are evidence in the economics literature that the sex of the eldest child matters as some papers have underlined a correlation and even a causation between the sex of the first born of a woman and her tolerance and exposure to IPV (Milazzo (2014)).

by the list. It could be consistent with evidence showing that the mother of a first born boy has stronger status within the household and may be less likely to experiment IPV as a result (Milazzo (2014)).

I also use a nonlinear least-squares estimation to account for the boundedness of the dependent variable. Table 3.7 shows the results of a logit estimation. I use the kict STATA package developed by Tsai (2019) that presents the output of the interactions between being treated with the list and the variables of interest as follows: for each type of IPV, the first column displays women's odds of experiencing IPV when they have a given characteristic, and the second column displays the correlation between a given characteristic and the predicted values of innocuous items that respondents answered "yes" to. The results obtained with the logit are consistent with those obtained with the linear estimation. Table 3.7 suggests that the odds of women residing in campaign clusters are $e^{1.199} \approx 3.32$ higher than the odds of women residing in noncampaign clusters. In addition, the odds that women older than 25 declare they experienced less severe physical violence are higher than the odds of women aged 18-25. I also find that having a first born daughter increases women's odds of declaring less severe physical violence by $e^{1.687} \approx 5.14$. Finally, the logit estimation suggests that muslim women are more likely to declare they experienced less severe physical violence than non-muslim women. As with the linear estimation, I do not find any source of heterogeneity in declaration of IPV with the LE for severe physical violence. Finally, the logit estimation suggests that the odds of women aged 25-35 to declare marital rape are $e^{-1.62} \approx 0.2$ lower than the odds of women aged 18 to 25. Overall, the results yielded by the logit estimation suggests than women aged 18 to 25, who are at the onset of their marital and fertile life, are more likely to declare they experienced IPV than women aged 25 and more. Estimating the impact of residing in a campaign cluster and of being the mother of a first born daughter on marital rape with a nonlinear function yields similar results as with the linear estimation.

Table 3.6: Heterogeneity in the prevalence of IPV under the LE

VARIABLES	(1) less severe	(2) less severe	(3) less severe	(4) less severe	(5) less severe	(6) less severe	(7) severe	(8) severe	(9) severe	(10) severe	(11) severe	(12) severe	(13) rape	(14) rape	(15) rape	(16) rape	(17) rape	(18) rape
treated with the list	0.172***	0.347***	0.204***	0.140**	0.206***	0.150***	0.190***	0.204*	0.186***	0.184***	0.189***	0.218***	0.161***	0.286**	0.236***	0.0941	0.210***	0.172***
	(0.0366)	(0.123)	(0.0503)	(0.0670)	(0.0510)	(0.0576)	(0.0387)	(0.111)	(0.0447)	(0.0618)	(0.0537)	(0.0583)	(0.0359)	(0.116)	(0.0410)	(0.0591)	(0.0479)	(0.0554)
Age category - ref.: 18-25																		
25-35		0.0738						-0.000662						0.0472				
		(0.128)						(0.123)						(0.123)				
35-45		0.105						-0.0420						0.0289				
		(0.116)						(0.107)						(0.127)				
45+		0.0213						-0.0114						0.0452				
		(0.119)						(0.136)						(0.123)				
treated with list x 25-35		-0.182						-0.0259						-0.0826				
		(0.141)						(0.143)						(0.137)				
treated with list x 35-45		-0.187						0.0208						-0.0687				
		(0.153)						(0.135)						(0.149)				
treated with list x 45+		-0.163						0.0269						-0.0985				
		(0.163)						(0.154)						(0.149)				
campaign clusters	0.135***						0.197***						-0.00780					
	(0.0406)						(0.0430)						(0.0399)					
treated with list x campaign clusters	0.0401						0.0275						0.0936*					
	(0.0492)						(0.0521)						(0.0484)					
some formal school			0.0501						-0.0525						0.0276			
			(0.0938)						(0.144)						(0.106)			
treated with list x formal school			-0.0622						0.103						-0.105			
			(0.0918)						(0.148)						(0.129)			
girl first child				0.0127						0.0503						-0.118		
				(0.0672)						(0.0703)						(0.0737)		
treated wth list x girl first child				0.103						0.0386						0.242**		
				(0.0843)						(0.0911)						(0.0938)		
mossi					-0.0915						-0.0118						-0.0288	
					(0.0975)						(0.115)						(0.0989)	
treated with list x mossi					-0.0215						0.0240						0.0131	
					(0.0798)						(0.0823)						(0.0730)	
muslim						-0.0986						-0.0268						-0.0639
						(0.0748)						(0.0996)						(0.0712)
treated with list x muslim						0.0660						-0.0226						0.0694
						(0.0824)						(0.0822)						(0.0760)
Observations	4,384	4,384	4,384	4,384	4,384	4,384	4,384	4,384	4,384	4,384	4,384	4,384	4,384	4,384	4,384	4,384	4,384	4,384
R-squared	0.032	0.246	0.245	0.245	0.245	0.245	0.032	0.234	0.234	0.234	0.234	0.234	0.024	0.181	0.181	0.186	0.181	0.181
mean dep var	1.342	1.342	1.342	1.342	1.342	1.342	1.042	1.042	1.042	1.042	1.042	1.042	1.261	1.261	1.261	1.261	1.261	1.261
pval test: var + list X var = 0	1.68e-08		0.873	0.0464	0.189	0.590	0		0.516	0.136	0.924	0.529	0.00487		0.412	0.0759	0.878	0.935
pval test: [25-35] + list X [25-35] = 0		0.204						0.780						0.701				
pval test: [35-45] + list X [35-45] = 0		0.367						0.820						0.619				
pval test: 45+ + list X 45+= 0		0.176						0.875						0.688				
pval of test robust to MHT (Simes method)	0.0032	NO	NO	NO	NO	NO	0.00	NO	NO	NO	NO	NO	0.0439	NO	NO	NO	NO	NO

Note: Dependent variable: number of marbles in a womn's left hand. Controls: strata fixed-effect (except for the regression with treated clusters as the independent variable), age, some formal education, muslim, mossi. Standard errors clustered at the strata level are in parenthesis.

Table 3.7: Correlates of IPV using the list experiment - logit

	less severe		severe		marital rape	
VARIABLES	LE	non-key items	LE	non-key items	LE	non-key items
campaign clusters	1.199**	0.201***	0.263	0.264**	0.618†	-0.0164
	(0.584)	(0.0681)	(0.476)	(0.104)	(0.383)	(0.0738)
Reference Age cat: 18-25						
25-35	-2.541***	0.294**	-0.511	0.0818	-1.620**	0.116
	(0.848)	(0.131)	(0.807)	(0.169)	(0.748)	(0.140)
35-45	-1.944**	0.297**	-0.142	0.00288	-0.741	0.0348
	(0.876)	(0.124)	(0.744)	(0.147)	(0.631)	(0.136)
45+	-2.750*	0.148	0.0542	0.0464	-1.285	0.106
	(1.464)	(0.137)	(0.852)	(0.195)	(0.830)	(0.140)
some formal schooling	-0.716	0.0472	0.564	-0.0904	-0.803	-0.0376
	(0.616)	(0.105)	(0.600)	(0.182)	(0.722)	(0.103)
eldest child is a girl	1.637**	-0.0433	0.421	0.0454	1.705***	-0.123
	(0.727)	(0.0785)	(0.569)	(0.101)	(0.618)	(0.0844)
muslim	1.091*	-0.100	-0.150	-0.144	0.447	-0.139*
	(0.584)	(0.0696)	(0.459)	(0.118)	(0.448)	(0.0794)
mossi	0.252	-0.183***	-0.0299	-0.0903	-0.392	0.0393
	(0.531)	(0.0657)	(0.492)	(0.103)	(0.405)	(0.0752)
polygamous	0.371	-0.0749	0.0684	0.186*	0.455	-0.0791
	(0.559)	(0.0654)	(0.564)	(0.106)	(0.486)	(0.0812)
Observations	4,384	4,384	4,384	4,384	4,384	4,384
pval chi-square	0.102	0.00112	0.986	0.0601	0.0560	0.247

Note The estimates displayed in the table are obtained using the kcit STATA package developed by Tsai (2019). The estimates in the column "LE" represents the correlates of IPV measured by the LE. They can be interpreted as odds ratios similarly to logistic regression coefficients. For example, the coefficient treated clusters in column "LE" suggests that, holding other variables constant, women's odds when residing in treated clusters of declaring less severe physical violence are $e^{1.119} \approx 3.32$ higher than the odds of experiencing less severe IPV than women residing in nontreated clusters. The estimates of column "nonkey items" are the predicted values of nonkey items that respondents answered "yes" to. Significance level: $***p < 0.01, **p < 0.05, *p < 0.1, \dagger p < 0.11$. Standard errors in parenthesis and clustered at the strata level.

3.6.4 Differences in declaration according to the method used

In this section, I first show the levels of prevalence of IPV with the two different methods among subgroups of respondents that I split according to women's characteristics. In a second step, I investigate whether women's characteristics correlate differently with IPV according to the method used to measure it.

Table 3.8: Prevalence with LE and direct questions by campaign clusters

	List Experiment	Direct	Diff	p-value	N
Panel A: campaign clusters					
less severe	24.65	17.04	7.61	.1868	2346
se	.057	.03			
severe	21.92	13.68	8.25	.1536	2346
se	.052	.026			
rape	25.65	10.18	15.48	.0014	2346
se	.052	.028			
Panel B: noncampaign clusters					
less severe	17.52	16.95	.57	.9109	2154
se	.047	.022			
severe	18.9	15.72	3.18	.5569	2154
se	.049	.027			
rape	15.99	13.72	2.27	.6457	2154
se	.041	.02			

Note: First column: output of a regression yielding the mean number of marbles in women's left hand in the treated group to the mean number of marbles in women's left hand in the control group. Controls: strata fixed-effect, age, some formal education, muslim, and mossi. Second column: output of regressing the direct question outcome on a constant. Standard errors clustered at the strata level.

I find descriptive evidence that there is a gap in women's reporting of IPV with the list and with the direct measure among women who reside in campaign clusters. Panel A of table 3.8 shows that the prevalence of marital rape with the LE is 15.56pp higher than the prevalence yielded with the direct measure. Panel B of table 3.8 underlines that there is no statistically significant difference among women who reside in noncampaign clusters.

I also find that women whose eldest child is a girl seem to declare more that they experienced less severe IPV and marital rape with the LE than they do when asked directly. The difference in declaration is of nearly 11pp for less severe physical violence, and of 20pp for marital rape. Panel A of table 3.9 shows there is no such difference for severe forms of physical violence. Panel B of table 3.9 suggests that unlike mothers of first born daughters, mothers of first born boys exhibit no significant difference in declaration of IPV according to the method used to measure it.

Finally, I find that the prevalence of less severe physical violence among women aged 18 to 25 years old is 31.16pp higher with the LE compared to the prevalence obtained with the direct

Table 3.9: Estimated prevalence with LE and direct questions by sex of the first born

	List Experiment	Direct	Diff	p-value	N
Panel A: Women whose first child is a girl					
less severe	24.97	14.05	10.92	.0722	2136
se	.051	.027			
severe	22.57	16.99	5.58	.4016	2136
se	.057	.032			
rape	31.04	11.34	19.7	.0011	2136
se	.062	.02			
Panel B: Women whose first child is a boy					
less severe	17.19	20.47	-3.29	.5963	2248
se	.065	.032			
severe	17.85	12.29	5.56	.3677	2248
se	.058	.019			
rape	9.34	12.24	-2.9	.619	2248
se	.055	.022			

Note: First column: output of a regression yielding the mean number of marbles in women's left hand in the treated group to the mean number of marbles in women's left hand in the control group. Controls: strata fixed-effect, age, some formal education, muslim, and mossi. Second column: output of regressing the direct question outcome on a constant.Standard errors clustered at the strata level.

measure, as shown in Panel A of table 3.10. Panel B of table 3.10 seems to suggest that among women older than 25, the prevalence of IPV using the LE is nearly 8pp higher than the one obtained when asking women directly. Tables A-3.10, A-3.11 and A-3.12 also describe patterns of heterogeneity according to the denomination, the ethnic group and the education level but are shown in the Appendix for the sake of brevity.

Table 3.10: Estimated prevalence with LE and direct questions by age

	List Experiment	Direct	Diff	p-value	N
Panel A: Women aged 18-25					
less severe	38.25	7.09	31.16	.0004	757
se	.086	.018			
severe	21.82	11.5	10.33	.3661	757
se	.103	.033			
rape	25.83	10.27	15.56	.1555	757
se	.106	.029			
Panel B: Women aged 25+					
less severe	17.45	18.82	-1.38	.7588	3743
se	.045	.022			
severe	21.71	15.16	6.54	.1292	3743
se	.04	.02			
rape	19.88	12.04	7.84	.0443	3743
se	.038	.02			

Note: First column: output of a regression yielding the mean number of marbles in women's left hand in the treated group to the mean number of marbles in women's left hand in the control group. Controls: strata fixed-effect, age, some formal education, muslim, and mossi. Second column: output of regressing the direct question outcome on a constant. Standard errors clustered at the strata level.

Table 3.11 and table 3.12 summarize how each variable of interest correlates with the LE measure (in column 1) and with the direct measure of IPV (in column 2). In the two tables, column 1 shows the coefficient of equation 3.4, in which I interact all the variables of interest with the dummy variable indicating that the respondent was exposed to the list with the sensitive item. Column 2 shows the coefficients of 3.5, and column 3 displays the pvalue of the difference between the coefficients in column 1 and column 2. Column 4 provides that pvalue adjusted for multiple hypothesis testing (MHT) using the Bonferroni method.

For less severe physical violence, table 3.11 shows that there two variables that correlates differently and significantly between the two measures: the sex of the eldest child of the respondent, and her age category at the time of the survey. I find that having a first born girl increases experience of less severe forms of physical violence by 18pp under the LE, whereas the direct measure suggests that it decreases experience of IPV by 6.5pp. I also find that the youngest age group, ie women aged 18 to 25, declare significantly more having experienced less severe IPV compared to women aged 25-35. Under the direct measure of IPV, the results are reversed and women aged 25-35 are 10pp more likely to declare they experience less severe forms of physical violence.

Table 3.12 shows that mothers of first born daughters are 22pp more likely to experience marital rape under the LE whereas the direct measure of IPV does not capture any significant effect of the sex of the first child. It also seems that women who reside in campaign clusters declare more IPV with the LE than they do with the direct measure. Even though this result is not robust to correcting for MHT, I provide a descriptive discussion in section B-3.1. of the pathways that may explain the difference between the two measures. I find no significant difference in declaration of IPV with the LE and with the direct measure for severe physical violence, as shown by table A-3.14 in Appendix. It is likely to be a byproduct of the floor effect detected for this list: it allows me to provide a valid, though underestimated, measure of severe physical violence, but it does not provide enough statistical power to compare sources of heterogeneity according to the method used.

Table 3.11: Correlates of IPV - less severe IPV

VARIABLES	(1) LE	(2) direct	(3) pval of diff.	(4) MHT adjusted p
treated with list	0.267***			
	(0.0837)			
campaign cluster	0.0328	0.0259	.93137596	1
	(0.0538)	(0.0196)		
ref cat age 18-25				
age 25-35	-0.295***	0.100***	.00061279	.00551507
	(0.0762)	(0.0274)		
age 35-45	-0.224***	0.0649**	.02646904	.23822133
	(0.0814)	(0.0295)		
age 45+	-0.161*	0.116***	.09586121	.8627509
	(0.0941)	(0.0335)		
some formal school	-0.0463	0.0169	.51541108	1
	(0.0656)	(0.0238)		
girl first child	0.181***	-0.0651***	.00318297	.02864671
	(0.0488)	(0.0179)		
muslim	0.0668	-0.0476**	.12861171	1
	(0.0580)	(0.0211)		
mossi	-0.0314	-0.0693***	.60159226	1
	(0.0564)	(0.0207)		
polygamous	0.0272	0.0203	.93248978	1
	(0.0528)	(0.0193)		
Observations	4,384	1,751		
R-squared	0.039	0.032		
mean dep var	1.342	0.178		
pval Joint significance test	0.0002	0.0000	.00345659	

Note In Column 1, the dependent variable is the number of items agreed upon by the respondent. Column 1 displays the coefficients on the interaction terms of equation 3.4 ie ρ_j. In Column 2, the dependent variable is the binary response to whether women experienced IPV when asked directly. Column 2 displays the coefficients δ of equation 3.5 for the control group only. Column 3 provides the pvalue of the difference between ρ_j and δ yielded by the suest command of STATA and column 4 shows the same value adjsted for multiple hypothesis testing using the Bonferroni method. The line "pval joint significance test" shows the pvalue of a joint significance test of the terms of interactions in column 1, the pvalue of the F-test in column 2 and the pvalue of the joint significance of the difference between ρ_j and δ in Column 3. Standard errors clustered at the strata level are in parenthesis.

Table 3.12: Correlates of IPV - marital rape

VARIABLES	(1) LE	(2) direct	(3) pval of diff.	(4) MHT adjusted p
treated with list	0.141*			
	(0.0422)			
campaign cluster	0.102*	-0.0187	.06391831	.57526475
	(0.0529)	(0.0168)		
ref cat age 18-25				
age 25-35	-0.150**	0.0367	.14745748	1
	(0.0749)	(0.0236)		
age 35-45	-0.0489	0.0154	.63119619	1
	(0.0800)	(0.0253)		
age 45+	-0.108	-0.0435	.64357892	1
	(0.0925)	(0.0288)		
some formal school	-0.0779	-0.0558***	.84052578	1
	(0.0645)	(0.0205)		
girl first child	0.220***	-0.0128	.00317278	.02855502
	(0.0480)	(0.0154)		
muslim	0.0315	-0.0809***	.17441366	1
	(0.0570)	(0.0182)		
mossi	-0.0601	-0.0410**	.78116645	1
	(0.0555)	(0.0178)		
polygamous	0.0575	0.00347	.55820125	1
	(0.0519)	(0.0166)		
Observations	4,384	1,751		
R-squared	0.030	0.037		
mean dep var	1.261	0.153		
pval Joint significance test	0.0001	0.0000	0.0017	

Note In Column 1, the dependent variable is the number of items agreed upon by the respondent. Column 1 displays the coefficients on the interaction terms of equation 3.4 ie ρ_j. In Column 2, the dependent variable is the binary response to whether women experienced IPV when asked directly. Column 2 displays the coefficients δ of equation 3.5 for the control group only. Column 3 provides the pvalue of the difference between ρ_j and δ yielded by the suest command of STATA and column 4 shows the same value adjsted for multiple hypothesis testing using the Bonferroni method. The line "pval joint significance test" shows the pvalue of a joint significance test of the terms of interactions in column 1, the pvalue of the F-test in column 2 and the pvalue of the joint significance of the difference between ρ_j and δ in Column 3. Standard errors clustered at the strata level are in parenthesis.

The type of LE implemented in this experiment is greedy in statistical power, which raises the following question: when I study the covariates of IPV using the LE, is the failure to reject the null hypothesis due to lack of impact of the covariate on IPV, or is it due to an insufficiently powered test? I first try to answer the question with a very raw approach by looking at the confidence interval (CI) of the estimates reproduced in Table 3.13. The CI provides the set of nonrefutable values of the parameter of interest. If the set of nonrefutable values is tightly clustered around a null-value, as suggested by Hoenig and Heisey (2001), we can be confident that the true value may be near null. Hoenig and Heisey (2001), however, does not provide a rule-of-thumb for the range of the CI, and as LE have high variance, it is not obvious what an acceptable range around 0 would be. Another suggestion by Hoenig and Heisey (2001) is to use equivalence testing instead of the usual hypothesis tests. The equivalence testing approach was initially developed in biopharmaceutics to determine whether two drugs have been shown to be equivalent. This type of tests reverses the burden of proof: instead of failing to show a difference between two drugs, one has to demonstrate that a large difference does not exist. Let's assume that one is willing to declare that the observed impact of a treatment, noted d, is negligible if the absolute value of the effect is below a given value Δ. Under the equivalence testing, the null hypothesis is $H_0 : |d| \geq \Delta$, i.e. the treatment has a large effect. The alternative hypothesis is $H_a : |d| < \Delta$, which is the hypothesis of practical equivalence (Hoenig and Heisey (2001)). In practice the equivalence hypothesis is tested using two one-sided tests (TOST) of level α as described in Schuirmann (1987). Schuirmann (1987) also shows that if the ordinary $1 - 2\alpha$ CI lies within $[-\Delta, \Delta]$, one can reject the null hypothesis of nonequivalence in favor of equivalence at the α level.[9] The rationale of this test can be applied to the study of the covariates of IPV with the LE. In table 3.12, I find that women belonging to the muslim group decreases report of marital rape with the direct question by 8pp, but I do not find any significant impact of belonging to the muslim group with the LE. Let's assume that I am only interested in the impact of belonging to the muslim group on IPV measured with the LE if it is at least twice as high as the one captured with the direct question, i.e. $2 * (-0.08) = 0.16$. The null hypothesis is then $H_0 : |d| \geq 0.16$, and $H_a : |d| < 0.16$. Table 3.13 shows the output of the equivalence test: with $\alpha = 0.05$, and assuming I am only interested in an effect twice as high as the one captured with the direct measure, $CI_{1-2\alpha}$ lies entirely within $[-\Delta_1, \Delta_1]$, which means I can reject the null hypothesis of nonequivalence in favor of equivalence at the 5% level. A different conclusion is reached if I set the Δ at 1.5, meaning that I would be interested in an effect of belonging to the muslim group 1.5 times the one captured with the direct measure of IPV. In this case, the interval $[-0.062, 0.125]$ does not completely lies in $[-0.1204, 0.1204]$,

[9]The version of equivalence testing presented here is deemed conservative by Hoenig and Heisey (2001).

and I fail to reject the null hypothesis of nonequivalence. For the binary variables indicating whether the respondent belongs to the mossi group, and the one indicating that respondent has some formal education, I find that $CI_{1-2\alpha}$ is not included in $[-\Delta, \Delta]$, whether I set Δ to twice the effect of the direct measure, or to 1.5 times the effect of the direct measure. It means that I fail to reject the null hypothesis that the impact of the mossi group and of education are not equivalent i.e. that they do not have 1.5 to twice the effect captured with the direct measure.

Based on this analysis, I argue that the inability to reject the null hypothesis that being a muslim has no impact on marital rape measured with the LE probably reflects an actual zero, whereas I do not argue it for belonging to the mossi group or having some formal school.

Table 3.13: Equivalence testing - Marital rape

Covariates	d_{direct}	d_{LE}	$CI_{1-\alpha}$ LE	$[-\Delta_1, \Delta_1]$	$[-\Delta_2, \Delta_2]$	$CI_{1-2\alpha}$
Muslim	-0.0809***	0.0315	[-0.0803, 0.143]	[-0.16, 0.16]	[-0.1204, 0.1204]	[-0.062, 0.125]
Mossi	-0.041**	-0.0601	[-0.1688, 0.0487]	[-0.082, 0.082]	[-0.0615, 0.0615]	[-0.151, 0.0312]
Some formal school	-0.0558***	-0.0779	[-0.2044, 0.0485]	[-0.112, 0.112]	[-0.084, 0.084]	[-0.184, 0.0282]

Note: The covariates chosen are the ones that have a significant impact on IPV measured with the direct measure but not with the LE. d_{direct} is the coefficient of the linear regression presented in column 2 table 3.12, d_{LE} is the coefficient of the linear regression presented in column 1 table 3.12, and $CI_{1-\alpha}$ in the confidence interval at the 5% level of the estimates yielded by the linear regression presented in column 2 table 3.12. $\alpha = 0.05$. The impact of the covariate on IPV measured with the LE that I assumed to be of interest in the equivalence testing are Δ_1 that is equal to twice the effect of the covariate on IPV measured with the direct question, and Δ_2 that is equal to 1.5 times the effect of the covariate on IPV measured with the direct question. $CI_{1-2\alpha}$ is the confidence interval computed following Schuirmann (1987) to carry out the test of equivalence. The null hypothesis of nonequivalence is rejected if $CI_{1-2\alpha}$ lies entirely between $-\Delta$ and Δ.

3.6.5 Does having a first born girl increase marital rape?

I suggest two potential pathways that could explain the relationship between IPV measured under the list and the sex of the first child. First, women whose first child is a daughter may be more at risk of experiencing marital rape. The literature has shown that in virilocal and patrolinear contexts, men are the pillars of the lineage and having boys is critical to assert women's footing in the household of her husband (Lesthaeghe (1989)). Second, the secrecy provided by the LE is expected to prompt women who are more vulnerable or more afraid to reveal IPV when asked directly to do so. Because their status in the household may not be as strong as mothers of a first born boy, mothers of a first born daughter may be more sensitive to the secrecy provided by the LE and report more IPV when asked with the LE. Let's call this pathway the report channel. Those pathways are not mutually exclusive.

In order to investigate the report channel, I use the test inspired by Aronow *et al.* (2015), which suggest that since the LE actually allows to circumvent any form of social desirability bias, then among respondents who are not afraid to reveal they experienced the sensitive behavior when asked directly, the declaration bias should not exist. According to them, a way to test this is to focus on the subsample of respondents who revealed they experienced the sensitive behavior when asked directly. Within this subsample, if people perfectly complied to the LE protocol, then the difference in the mean number of answers people agree with in (T) and (C) should tend towards 1. I use their rationale to investigate whether, among women who declared IPV under the list, respondents residing in campaign clusters and in noncampaign clusters comply in different ways to the LE protocol.

The evidence provided here is essentially descriptive as zooming in on smaller samples leads to a decrease in statistical power. In addition, as Tsai (2017) underlines it, the method suggested by Aronow *et al.* (2015) is for univariate analysis and relies on the assumption that the direct self-report on the sensitive item is only prone to over-reporting ($Y_i = 1, L_i = 0$) or under-reporting ($Y_i = 0, L_i = 1$) but not both, which Tsai (2019) describes as a major limitation of the method and which seems to be the case in this data.[10] Despite these limitations, I believe this method is useful for descriptive purposes and still allows to gather a body of corroborating evidence for one hypothesis or the other.

First, I zoom in on women who declared they experienced IPV under the direct measure. Table A-3.13 suggests that for marital rape, women seem to comply in a similar manner regard-

[10] See section 3.6.6 for details.

less of the sex of their eldest child, unlike what was noted for women residing in noncampaign clusters. As a result, the reporting channel does not seem relevant here.

Second, I focus on women who declared they did not experience marital rape when asked directly, and estimate how many of them did so with the list experiment. Estimates in panel A and B of table 3.14 underline that among mothers of a first born girl who declared no marital rape when asked directly, 28% switched their answer to reveal they experienced marital rape under the list. Mothers of first born boys do not exhibit such pattern, suggesting that mothers of a first born girl are more likely to experience marital rape than mothers of a first born boy. It is interesting to note that for severe violence, the pattern seem to be reversed (though the difference in the share of "switchers" between panel A and panel B is not statistically significant); mother of first born boys appear more likely to use the list to declare IPV than mothers of first born daughters. Let's remember that the list used to measure severe IPV suffered from a floor effect, which tends to protect less than the other lists the plausible deniability of women if they choose to reveal IPV. Despite the lesser protection provided by this list, mothers of first born boys still used the list to reveal their experience of severe physical violence under the LE whereas it seems to offer too weak a protection for mothers of first born daughters to do so.

Table 3.14: LE prevalence of women who did not reveal IPV with the direct measure by sex of first child

	less severe	severe	rape
Panel A: Women who declared no IPV with the direct measure - first child girl			
Treated with list			
	0.153**	0.087	0.281***
	(0.07)	(0.08)	(0.08)
Mean Dep. Var.	1.27	1.00	1.15
N	1694	1807	1758
r2	0.42	0.40	0.36
Panel B: Women who declared no IPV with the direct measure - first child boy			
Treated with list			
	0.019	0.149*	-0.007
	(0.08)	(0.08)	(0.06)
Mean Dep. Var.	1.32	0.97	1.23
pval test equal share of switchers between Panel A and Panel B	0.14	0.55	0.001
N	1795	1927	1866
r2	0.37	0.36	0.32

Note: The dependent variables are the number of items women agreed with in the LE. I include strata fixed-effect and control for women's age category, whether she belongs to main ethnic group (mossi) and whether she belongs to the main religious group (muslim). Standard errors clustered at the strata level are in parenthesis. Significance level: $***p < 0.01, **p < 0.05, *p < 0.1$.

3.6.6 Adjusting IPV prevalence to non-compliance with the protocol

As mentionned in section 3.6, the data suggests that some women did not reveal their experience of IPV with the list but seem to have done so when asked directly. In order to formally assess the magnitude of this behavior, I use a test *a la* Aronow *et al.* (2015) to investigate potential complex behavioral responses of the respondents to the experiment and, when relevant, adjust the estimates to these behaviors.

Table 3.15: Estimated prevalence with LE among respondents who said "yes" to the direct question

VARIABLES	(1) less severe	(2) severe	(3) rape
treated with list	0.686***	0.529***	0.640***
	(0.203)	(0.177)	(0.212)
Observations	915	665	781
R-squared	0.653	0.703	0.672
Controls	Yes	Yes	Yes
pval estimates=1	0.122	0.00816	0.0907

Note: Controls: strata fixed-effect, age, some formal education, muslim, mossi, ipv at baseline. Standard errors clustered at the strata level in parenthesis. Significance level: $***p < 0.01, **p < 0.05, *p < 0.1$.

As shown in table 3.15, I find that the difference in means of the number of sentences women agree with does not tend towards 1 in the subsample of women who answered "yes" to IPV when asked directly about severe violence and marital rape.[11] In other words, not all women who declare they experienced IPV when asked directly, took the opportunity of declaring their experience of IPV with the list.The non-compliance is unlikely to be driven by women who did not understanding the LE protocol. Indeed, enumerators thoroughly explained and showed the respondents how the experiment worked before they were asked to do it. In addition, a dummy list was used prior to the actual lists to make sure that the respondent understood what to do. The enumerators were told to clarify the instructions and redo the dummy list if women failed to understand the protocol the first time. Another explanation would be that the plausible deniability offered by the LE protocol is not enough to encourage some women to reveal they were the victim of IPV the first time they are asked. The fact that some women belong to the list-conceal/direct-reveal category finds meaning if we turn to the work of sociologists such as Smith (1994) that underlined that for researchers to reach a more accurate measure of

[11] For severe violence, the non-compliance seem larger than for marital rape. It is a likely reflection of the floor effect mentionned earlier for this outcome.

IPV, women should be given several opportunities to reveal IPV by asking them, in different ways, whether they experienced interpersonal violence. According to Smith (1994), it is a matter of building trust with the respondent around the topic of IPV. The results in table 3.15 thus suggest that when women are asked directly about IPV about a precise type of violence once in household surveys, we underestimate the actual direct prevalence of violence.

Based on the framework suggested in section 3.5, women who revealed IPV with the direct measure but did not do so with the LE are found along the red dashed line of figure 3.4. They should, however, be found along the blue dotted line of figure 3.4. For instance, among women who declared having experienced marital rape when asked directly, $100-64\% = 36\%$ of women who received the baseline list + the sensitive item through the list are "missing" in the sense that they did not declare IPV under the list. A back-of-the enveloppe calculation suggests that I am "missing" $0.36 * 0.6643 * 781 \approx 187$ women.[12] If these 187 women had revealed they experienced marital rape under the list, I would have had a prevalence of marital rape of $[\frac{(0.2113*4500)+187}{((0.2113*4500)+187)+(0.7887*4500)}] * 100 \approx 24\%$.

3.6.7 Adjusting direct IPV prevalence to priming

Table 3.16: Estimates of IPV with direct questioning in (T) and(C)

VARIABLES	(1) less_severe	(2) severe	(3) rape
Treated with list	0.0329	0.00575	0.0412*
	(0.0226)	(0.0216)	(0.0239)
Observations	4,500	4,500	4,500
R-squared	0.249	0.258	0.212
Controls	Yes	Yes	Yes
Mean control	0.170	0.146	0.118

Note: Controls: strata fixed-effect, age, some formal education, muslim, mossi. Standard errors in parenthesis and clustered at the strata level. Significance level: $***p < 0.01, **p < 0.05, *p < 0.1$.

In addition to women belonging to the list-conceal/direct-reveal category, table 3.16 shows that being exposed to the sensitive item in the LE before the direct question about IPV, increases women's likelihood to self-report marital rape by 4pp. It underlines that prior exposure to the sensitive item alters women's likelihood to reveal IPV under the direct measure. If all women

[12] 0.66 is the share of women among those who answered "yes" to the direct measure of marital rape, who were exposed to the baseline list + the sensitive item and 781 is the number of women who ackowledged to have experienced marital rape when asked directly.

were exposed to the sensitive item about marital rape before being asked directly about it, the prevalence would be $[\frac{(0.12+0.04)*4500}{((0.12+0.04)*4500)+(0.84*4500)}] * 100 = 16\%$, instead of 12% if none of the women were primed, which is closer to the 10% prevalence measured with the DHS for Burkina Faso in 2010 among women aged 18 to 49 years old in a union at the time of the survey.[13] The priming behavior exhibited by $0.04 * 0.66 * 4500 \approx 119$ respondents could explain up to 62% of the non-compliance underlined in table 3.15.

3.6.8 Other heterogeneity

I have shown that some women are sensitive to being exposed twice to the sensitive item: a first through the LE, and a second time by being asked the direct question on IPV. I now explore whether there are salient patterns in this behavioral response of women. I examine whether the priming behavior correlates with the following variables: whether the woman lives in a campaign cluster, whether her first child child is a girl, the education level, the age category, whether women are married to a polygamous partner, the ethno-linguistic group, and the religious group.

Table 3.17: Heterogeneity in priming patterns under the direct measure of IPV

VARIABLES	(1) less severe	(2) severe	(3) rape
treated with list x campaign clusters	-0.0356	0.0229	0.0141
	(0.0388)	(0.0389)	(0.0336)
ref.: age category 18-25			
treated with list x 25-35	-0.0440	-0.0142	0.0130
	(0.0512)	(0.0585)	(0.0524)
treated with list x 35-45	0.0418	0.0906*	0.0623
	(0.0495)	(0.0528)	(0.0559)
treated with list x 45+	-0.00189	0.0103	0.0492
	(0.0603)	(0.0715)	(0.0506)
treated with list x some formal schooling	0.0646	-0.00956	0.114**
	(0.0744)	(0.0769)	(0.0546)
treated with list x girl first child	0.127**	-0.00360	0.0192
	(0.0490)	(0.0411)	(0.0305)
treated with list x muslim	-0.00118	0.0131	0.0129
	(0.0387)	(0.0441)	(0.0316)
treated with list x mossi	-0.0165	-0.0233	0.00899
	(0.0407)	(0.0413)	(0.0350)
treated with list x polygamous	0.0307	0.0447	0.0562
	(0.0493)	(0.0470)	(0.0356)
Observations	4,384	4,384	4,384
R-squared	0.038	0.035	0.029
mean dep var	0.178	0.138	0.153
pval joint sign. interactions	0.422	0.569	0.0578

Note: Controls: strata fixed-effect, age, some formal education, muslim, mossi. Standard errors in parenthesis and clustered at the strata level. Significance level: $***p < 0.01, **p < 0.05, *p < 0.1$.

[13] Author's calculation.

I interact all the variables with the exposure to the sensitive item with the list and show the interaction terms in table 3.17. Based on the last line of the table, the interaction terms are jointly significant for marital rape only, so I will focus on this dependent variable only. The estimates in the third column of table 3.17 underline that women with some formal schooling are more likely to report they experienced marital rape under the direct measure after they were exposed to the sensitive item in the LE. The educated women who exhibit a priming behavior could either have used the LE to reveal IPV; in this case, they would reveal IPV under the list and keep doing so with the direct measure, or they did not disclose their experience of marital rape with the LE, but they do it under the direct measure, when asked a second time. To get a sense of whether educated women belong to one category or the other, I split the sample according to whether women have some formal education and (i) examine whether women who revealed IPV under the direct-measure also did it under the LE using again the test inspired by Aronow *et al.* (2015), (ii) investigate whether women who declared no IPV under the direct measure are more likely to have switched answer and revealed their experience of IPV under the LE.

Panel A and panel B of table 3.18 present the output of the test inspired by Aronow *et al.* (2015) on the subsamples of women who declared they experienced marital rape with the direct measure, with and without formal education respectively. Panel C provides the share of educated women who revealed IPV with the list but did not do it with the direct measure (the "switchers"). Panel D shows the same output among women with no formal education. Though I lack statistical power to identify significant patterns among women with some formal education in panel A, panel B seem to suggest that women with no formal education who experienced IPV and revealed it under the direct measure, also revealed it under the LE.[14] If we keep focusing on marital rape, panel D underlines that 16% of women with no formal education who said they did not experienced marital rape when asked directly, used the list experiment to reveal they were the victim of such an abuse.[15]

To summarize, women with no formal education who declared experiencing marital rape when asked directly are, by and large, also likely to have done so with the list. Yet, among those who deny that there were the victim of marital rape when asked directly, 16% took the opportunity of the list to say otherwise, which suggests that they switched answer with the direct measure, when they are not protected by the plausible deniability of the LE. It seems that women with

[14]For women without formal education, the coefficient is not significantly different than 1. The lack of statistical power does not allow to draw conclusions about women with some formal education who declared they experienced marital rape when asked directly.

[15]Though panel C seem to suggest that, overall, women with some formal education who did not reveal IPV under the direct measure, did not reveal IPV under the list, the estimation is too noisy to allow a definitive conclusion.

some formal education react in a different way to being exposed twice to the sensitive item about marital rape. There are shreds of evidence suggesting that they do not switch their answer from one period to the other, but rather provide a more consistent pattern of answers with the two measures: they reveal marital rape with the LE and keep doing so with the direct measure. The data supports the idea that prior exposure to the sensitive item with the list increases educated women's probability to reveal IPV under the direct measure. It is as if the answer to the list acted as a commitment device in revealing experience of marital rape when asked directly. Women with some formal education are no less reluctant than uneducated women to reveal they were the victim of IPV, but they might feel bound by the answer they gave under the list experiment.

Table 3.18: Heterogeneous response with the LE by education level

	less severe	severe	rape
Panel A: Direct-reveal with some formal education			
Treated with list			
	0.082	-0.051	0.959
	(0.60)	(1.35)	(0.85)
Mean Dep. Var.	1.11	0.59	1.15
pval H0=1	0.13	0.44	0.96
N	149	108	122
r2	0.95	0.98	0.95
Panel B: Direct reveal with no formal education			
Treated with list			
	0.597***	0.630***	0.721***
	(0.23)	(0.19)	(0.27)
Mean Dep. Var.	1.45	1.11	1.42
pval H0=1	0.08	0.05	0.31
N	766	557	659
r2	0.67	0.66	0.69
Panel C: Women who declared no IPV under the direct measure with formal education			
Treated with list			
	0.117	0.129	-0.131
	(0.21)	(0.22)	(0.19)
Mean Dep. Var.	1.37	1.02	1.19
N	539	580	566
r2	0.61	0.57	0.56
Panel D: Women who declared no IPV under the direct measure with no formal education			
Treated with list			
	0.128**	0.137**	0.167***
	(0.06)	(0.05)	(0.04)
Mean Dep. Var.	1.27	0.98	1.19
N	3046	3255	3153
r2	0.30	0.29	0.25

Note: The dependent variables are the number of items women agreed with in the LE. I include strata fixed-effect and control for women's age category, whether she belongs to main ethnic group (mossi) and whether she belongs to the main religious group (muslim). Standard errors clustered at the strata level are in parenthesis. Significance level: $***p < 0.01, **p < 0.05, *p < 0.1$.

3.7 Discussion and Methodological consideration

3.7.1 Motives of under-declaration: self-image and social pressure

I have shown that providing plausible deniability to women makes them more likely to reveal they experienced IPV. As the LE conceals their answer, women may less fearful of the physical and emotional retaliation their husband could exert, should they become aware that their wife revealed she was abused. However, all surveys that include questions about IPV require that women are isolated from their surroundings when being asked such questions. As a result, the fear of retaliation is probably not the only driver of the discrepancy between direct and indirect measures of IPV.

Social desirability bias is the main other motive one can think of and the one that the LE is meant to tackle. As mentionned in Section 3, the social desirability bias at play here could originate in social pressure: women may be afraid to be judged as disloyal to their husband by other members of the household or by their neighbours or even by the enumerator.[16] This hypothesis is particularly likely to hold in the tight-knit societies of rural Burkina Faso. The social desirability bias could also originate in self-perception or self-image issues, as women may feel shame in declaring they were the victim of the husband's violence for instance.

I rely on the framework of Bursztyn and Jensen (2017) to factor in social image concerns in the determinants of women's decision to reveal they experienced IPV. When a respondent reveals she experienced IPV, she provides information about whether she belongs to the socially desirable type (for instance, the wife who protects the image of her husband) or to the socially undesirable type. The respondent expects benefits from being perceived as being of the socially desirable type by a reference group (the enumerator or the neighbors for instance in this case). Individuals in the reference group determine whether the respondent is of the socially desirable type, conditional on observing her choice of action (here, revealing IPV or not). In Bursztyn and Jensen (2017), the social image concern is a product of the respondent expected benefits from being perceived of the socially desirable type, of the probability that individuals in the reference group think that she is of the socially desirable type conditional on observing her choice of action, and of how much she cares about being perceived by the reference group as of the socially desirable type. The desirability of revealing that one was the victim of IPV is then a function of social image, and of benefits and costs unrelated to image concerns.

[16] In Section 3, I also underline that women may be afraid or shameful to be suspected to have been punished by their husband because it would suggest they did something wrong also belongs to the social pressure motive.

By weakening the observability of women's answer to the question on IPV, the LE impacts the probability that individuals in the reference group think that she is of the socially undesirable type, in a larger extent than the direct question does, which decreases the discomfort of revealing that one was the victim of IPV. Even though the design of the experiment does not allow to disentangle the effect of self-image from that of social pressure, some insights may be found in other papers using alternative methods to measure IPV. Cullen (2020) finds that the ACASI method, where the respondent self-administer the question about IPV using a pre-recorded message, yields estimates that are not significantly different than the one yielded by the direct measure of IPV but that are significantly lower than the one obtained with the list experiment among women in Rwanda. Even though Cullen (2020) does not formalize nor phrase it this way, since the ACASI method increases the privacy of IPV declaration but does not weaken the signal of declaring IPV, it is consistent with the hypothesis that self-image might not be relevant factor in the under-declaration of IPV among women.

3.7.2 Methodological considerations

Despite its growing popularity, the analysis of data yielded by the list experiment is not straightforward (see Tsai (2019) for a summary of the different methods used and their limitations). For now, the literature mainly relies on linear estimations as it is the most convenient approach. Imai (2011) and Blair and Imai (2012), however, recommend to use a nonlinear least squares estimation to account for the boundedness of the dependent variable. Imai (2011) also developed an estimator that models the joint probability of the non-sensitive items and the sensitive items using a maximum likelihood estimator, to improve the efficiency of the statistical estimation. Tsai (2019) nevertheless underlines that the estimator developed by Imai (2011) assumes that the count of innocuous items follows a binomial law, which implies that (i) each respondent's answer to an innocuous item is independent of the answer given to any other innocuous items, and (ii) that the innocuous items have an equal probability of success (i.e. answering positively to them). However, by construction, it is not the case in this work nor in the literature using list experiments, because it goes against practitioner's advice to limit the variance in the data; Glynn (2013) notably suggests to select innocuous items that are negatively correlated. In addition, selecting an innocuous item with low probability and another one with high probability - which is the choice is this study - is an obvious and convenient way of avoiding floor or ceiling effects. Finally, in practice, having independent and equally distributed innocuous items would require intensive and costly piloting of the innocuous items. As a result, the estimator developed by Imai (2011) does not fit the profile of my data. The question of how to model the

count of the innocuous items in a way that makes estimators relevant for practitioners is a key and well-identified limit of the econometric literature on the subject (Tsai (2019)).

In order to overcome this issue, Corstange (2009) suggested an alternative list experiment design where instead of using a list to obtain the number of innocuous items the respondent agrees with, respondents in the control group are directly asked to answer "yes" or "no" to the innocuous items.[17] Corstange (2009) then uses the binary information for each innocuous items to model the marginal probabilities of the sensitive item (L_i) and each of the innocuous items simultaneously.[18] I use the binary answer to each of the non-sensitive item in the super control group to use the estimator developed by Corstange (2009).[19]

Finally, Tsai (2017) builds on the work of Aronow *et al.* (2015) and Eady (2017) to combine the measure of the sensitive item with the list experiment and its measure with the direct question to build an estimator meant to increase the statistical efficiency of the list experiment estimation. The estimator of Tsai (2017) also alleviates the constraint that the direct self-report of the sensitive item be only prone to over-reporting ($Y_i = 1, L_i = 0$) or under-reporting ($Y_i = 0, L_i = 1$) but not both.[20]

In table 3.19, I compare the results yielded by four different estimators to assess the relation between a set of respondent's characteristics and their probability to declare IPV using the list experiment. The four estimators are as follows: linear, non-linear, the maximum likelihood estimator of Imai-Tsai-Eady that makes use of the direct answer to the sensitive item in the estimation and the maximum likelihood estimator of Corstange-Imai. I use Tsai (2019) kict STATA package to perform the estimations. As mentionned in Section 3.6.1, the results obtained for marital rape with OLS and with the non-linear estimation are overall consistent as I find that residing in a campaign cluster and having a first born daughter lead to more declaration of

[17] The treatment group receives the sensitive item within a list and provides the total number of items they agree with as in the original design.

[18] Whether or not it is equivalent in practice to ask the total count of innocuous items a respondent agrees with and to sum the binary response to each innocuous items asked directly is debated. First, it relies on somehow more stringent assumptions that the ones of the regular design: respondents have to provide a truthful answer to the sensitive item and the treatment status should not influence answers to each of the non-sensitive questions (rather than the simple no-design effect of the regular version of the list experiment). Empirically, unlike Flavin and Keane (2009), Tsuchiya *et al.* (2007) find that the sum of the non-sensitive items is systematically inferior to the count in the control group who gets the baseline list, which leads to an overestimation of the sensitive behavior measured with the list experiment. In this work, I compare the sum of binary answers to the innocuous items in the super control group and the total count of innocuous items respondents agree with in the control group of the standard list experiment design and find no statistically significant difference between the two.

[19] I actually use the version of the Corstange (2009) improved by Imai (2011).

[20] A limitation that was attached to Aronow *et al.* (2015) and Eady (2017).

marital rape. The detrimental impact of having a first born daughter on declaration of marital rape seems robust across the four estimators. The estimator of Imai-Tsai-Eady also captures a negative association between belonging to the mossi ethno-linguistic group (which is the main ethno-linguistic group in Burkina Faso) and declaring IPV with the LE. As for the maximum likelihood estimator of Corstance and Imai, it suggests a negative association between declaring marital rape and having some formal schooling, and consistently with the nonlinear estimator, it also underlines that women's declaration of marital rape is positively associated with belonging to the younger age-group (18-25) compared to women being 25 to 35 at the time of the survey.

Table 3.19: Correlates of marital rape using the list experiment using different estimators

VARIABLES	OLS	Logit	ML Imai-Tsai-Eady	ML Corstange-Imai
some formal schooling	-0.0779	-0.803	-0.210	-0.986**
	(0.102)	(0.722)	(0.330)	(0.402)
mossi	-0.0601	-0.392	-0.834***	0.103
	(0.0636)	(0.405)	(0.322)	(0.340)
muslim	0.0315	0.447	0.706*	-0.0353
	(0.0750)	(0.448)	(0.391)	(0.330)
Reference Age category : 18-25				
25-35	-0.150	-1.620**	-0.0944	-0.864**
	(0.115)	(0.748)	(0.456)	(0.420)
35-45	-0.0489	-0.741	0.588	-0.894*
	(0.126)	(0.631)	(0.448)	(0.540)
45+	-0.108	-1.285	-0.183	-0.636
	(0.128)	(0.830)	(0.495)	(0.486)
campaign clusters	0.102*	0.618†	-0.0797	-0.0375
	(0.0608)	(0.383)	(0.297)	(0.354)
eldest child is a girl	0.220***	1.705***	1.077***	0.832**
	(0.0750)	(0.618)	(0.336)	(0.336)
polygamous	0.0575	0.455	0.0692	0.290
	(0.0889)	(0.486)	(0.294)	(0.431)
Observations	4,384	4,384	4,384	3,035

Note The estimates displayed in the table are obtained using the kcit STATA package developed by Tsai (2019). The estimates in the column ″LE″ represents the correlates of IPV measured by the LE. They can be interpreted as odds ratios similarly to logistic regression coefficients. For example, the coefficient treated clusters in column ″logit″ suggests that, holding other variables constant, women's odds of declaring marital rape when residing in campaign clusters are $e^{0.618} \approx 1.85$ higher than the odds of experiencing marital rape of women residing in noncampaign clusters. Significance level:$***p < 0.01, **p < 0.05, *p < 0.1, \dagger p < 0.11$. Standard errors in parenthesis and clustered at the strata level.

3.7.3 Using LE in household surveys to measure IPV?

The LE could be a powerful tool to provide a less biased measure of the overall prevalence of IPV among populations surveyed during large household surveys like the DHS. More research is needed however, to understand the heterogeneous response of women to using this survey tool to measure IPV. It may be difficult though as the LE is a method that has high variance, which may prevent researchers from investigating heterogenous effects among subgroups of respondents for lack of statistical power. The version of the LE I implemented in this paper is greedy in statistical power but this pitfall could be limited by using double list experiments (DLE) (Glynn (2013)), and by stratifying on variables of interest. Adding a LE to household surveys is a low cost survey tool but, in order to collect meaningful estimates of IPV with the LE, it will require more time and precautions than the direct questions about IPV. First, as it is recommended by several guidelines on data collection on IPV, the enumerators should be of the same gender as the respondent, and speak the same language. Second, training enumerators to perform the protocol of the LE is pivotal to the success of the method, and it requires an intensive training, both theoretical and practical. The LE is costly in times in the sense that it lenghtens significantly the duration of the survey; enumerators have to explain the protocol, perform a dummy list and then the mere fact of going through the lists takes longer than simply adding the direct questions. It may lead to a trade-off between obtaining less biased estimates of the types of violence that the researcher chooses to measure, and the number of different types of IPV that the research can investigate. Another limitation of the LE is that it can be cognitively taxing for the respondent. Even though there is no literature that assesses the optimal number of LE that should be presented to a respondent, it is important to keep in mind that the respondent may get weary at being presented too many LE. The LE also requires to pilot rather intensively the innocuous items of the LE, to make sure that they are not correlated in any way with the sensitive item, but also to try and have a sense of their frequency. Having a sense of their frequency is important to avoid floor or ceiling effects but also to anticipate whether the researcher will be able to use the tools developed by econometricians and discussed in a section 3.7.2. Given the current state of the knowledge on measures of IPV, I would recommend to add LE to household surveys if and only if a thorough training of the enumerators and a detailed piloting of the LE can be performed. In my opinion, even though the direct measure of IPV is known to be flawed, the current knowledge on using LE to measure IPV is at too early a stage to completely substitute it for direct measures in surveys.

3.8 Conclusion

I show that the direct measure of IPV underestimates the occurence of severe physical violence and marital rape by 7 to 9 percentage points. Unlike what is conveyed by direct measures of IPV, women seem to equally suffer from less severe and severe physical violence and marital rape. I also find suggestive evidence of a preference for boys that shows through an acute exposure to marital rape under the LE measure, whereas it is not detected with the direct measure. I also find descriptive evidence that women exposed to a mass media program meant to encourage modern contraception uptake seem to be empowered enough by the radio program to seize the opportunity of the LE to declare they were the victim of marital rape. Thanks to safeguard mechanisms included in the design of the LE, I am also able to explore the consistency the LE where much of the literature does not. Even though the LE reduces bias in declaration, it does not circumvent all the mechanisms that lead to underestimate experience of IPV. As a result, LE estimates should be read as a low bound for IPV, and direct measures of IPV as the lowest. Finally, I also use this work to test the tools developed by econometricians to improve the statistical exploitation of the data yielded by list experiments. The sophistication of the tools built by econometricians significantly improved the information that can be derived from list experiments, but a continued dialogue with practitioners may be needed to bring these tools and practices on the field closer to one another.

Appendix A

Table A-3.1: No-design effect - Less severe IPV

Less severe physical violence	***Number of items applying to respondent's situation***				
	0	*1*	*2*	*3*	*4*
Treatment list - proportion at least	1	0.915	0.478	0.15	0.026
Control list - proportion at least	1	0.914	0.379	0.06	0
Row1 - Row2	.	0.001	0.099	0.09	0.026

Note: N = 4500

Table A-3.2: No-design effect - Severe IPV

Severe physical violence	***Number of items applying to respondent's situation***				
	0	*1*	*2*	*3*	*4*
Treatment list - proportion at least	1	0.796	0.358	0.1	0.015
Control list - proportion at least	1	0.761	0.258	0.029	0
Row1 - Row2	.	0.035	0.1	0.071	0.015

Note: N = 4500

Table A-3.3: No-design effect - Marital rape

Marital rape	***Number of items applying to respondent's situation***				
	0	*1*	*2*	*3*	*4*
Treatment list - proportion at least	1	0.92	0.461	0.121	0.017
Control list - proportion at least	1	0.904	0.331	0.041	0
Row1 - Row2	.	0.016	0.13	0.08	0.017

Note: N = 4500

Table A-3.4: Balancing test

	Control	Treated	p-value	Obs
age	34.232	33.74	0.364	4500
muslim	0.648	0.65	0.591	4500
mossi	0.593	0.6	0.704	4500
no formal education	0.816	0.82	0.783	4500
some formal education	0.184	0.18	0.783	4500
no physical ipv at baseline	0.846	0.84	0.692	4500
physical ipv at baseline	0.152	0.16	0.764	4500
no info on physical ipv at baseline	0.002	0	0.263	4500
no marital rape at baseline	0.94	0.94	0.884	4500
marital rape at baseline	0.058	0.06	0.998	4500
no info on marital rape at baseline	0.002	0	0.263	4500
husband education	0.266	0.24	0.238	4368
polygamous partner	0.518	0.54	0.533	4500
age at first union	17.727	17.76	0.813	4109
age first child	19.115	19.26	0.303	4123
sex first child	0.49	0.51	0.989	4384
respondent earns money	0.485	0.48	0.753	4500

Table A-3.5: Correlation of non-key items

list less severe physical violence			
	meals alone	*funeral*	*dom chore*
meals alone	1.0000		
funeral	-0.0213	1.0000	
dom chore	0.0968	0.0213	1.0000
list severe physical violence			
	partner sweeps courtyard	*go to the market*	*sick partner*
partner sweeps courtyard	1.0000		
go to the market	0.0787	1.0000	
sick partner	0.0906	0.1925*	1.0000
list marital rape			
	help with husband's laundry	*partner meal*	*bought new wax*
help with husband's laundry	1.0000		
partner meal	-0.0254	1.0000	
bought new wax	0.0595	0.1144	1.0000

Note: N = 212

Table A-3.6: No-design effect - Less severe IPV

	Less Severe IPV			
	$\hat{\pi_y}0$		$\hat{\pi_y}1$	
Response value	*Proportion*	*SE*	*Proportion*	*SE*
0	0.0853	0.0049	0.001	0.0083
1	0.4369	0.011	0.0979	0.0146
2	0.2301	0.0132	0.089	0.0084
3	0.0359	0.0063	0.0239	0.0027

Note: the table shows for a given sensitive item, the estimated proportion of respondent characterized by the total number of affirmative answers to the control questions, y, and the truthful answer for the sensitive item, z (1 indicates affirmative and 0 represents negative). Following Blair and Imai (2012), this proportion is noted $\hat{\pi_y}z$.

Table A-3.7: No-design effect - Severe IPV

	Severe IPV			
	$\hat{\pi_y}0$		$\hat{\pi_y}1$	
Response value	*Proportion*	*SE*	*Proportion*	*SE*
0	0.2041	0.0071	0.0346	0.0124
1	0.4045	0.0132	0.0992	0.0134
2	0.1588	0.0117	0.0701	0.0066
3	0.0156	0.0045	0.0132	0.002

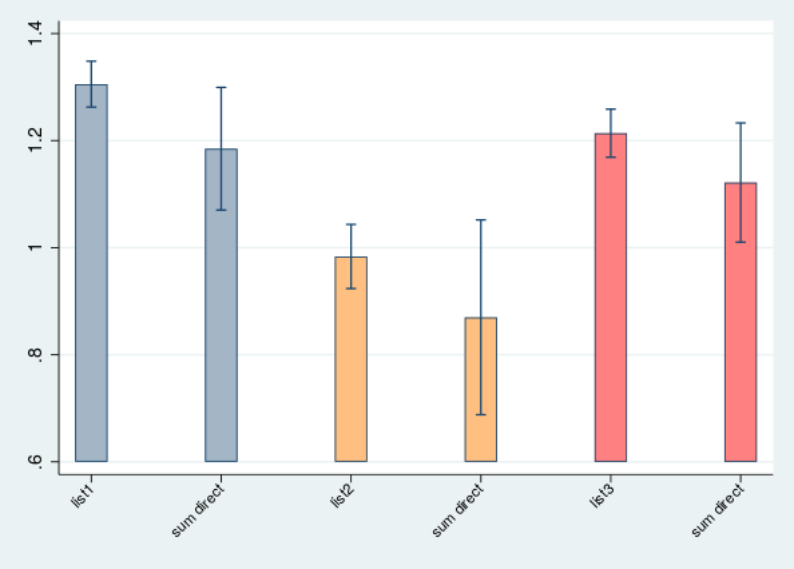
Note: the table shows for a given sensitive item, the estimated proportion of respondent characterized by the total number of affirmative answers to the control questions, y, and the truthful answer for the sensitive item, z (1 indicates affirmative and 0 represents negative). Following Blair and Imai (2012), this proportion is noted $\hat{\pi_y}z$.

Table A-3.8: No-design effect - Conjugal Rape

	Conjugal Rape			
	$\hat{\pi_y}0$		$\hat{\pi_y}1$	
Response value	*Proportion*	*SE*	*Proportion*	*SE*
0	0.0798	0.0047	0.0163	0.0085
1	0.4437	0.0112	0.1296	0.0143
2	0.2113	0.0126	0.0785	0.0074
3	0.0252	0.0052	0.0156	0.0022

Note: the table shows for a given sensitive item, the estimated proportion of respondent characterized by the total number of affirmative answers to the control questions, y, and the truthful answer for the sensitive item, z (1 indicates affirmative and 0 represents negative). Following Blair and Imai (2012), this proportion is noted $\hat{\pi_y}z$.

Figure A-3.1: Mean number of innoucuous items agreed upon with the LE and when asked directly



Note: Sample: Control groups only.

Table A-3.9: Estimated prevalence of IPV with LE

	T1C1	T2C2	Diff	p-value
less severe	23.1	20.72	-2.38	0.7219
se	0.048	0.057		
severe	22.11	17.03	-5.08	0.5456
se	0.055	0.056		
rape	21.06	22.03	0.97	0.9085
se	0.047	0.064		
N	2249	2251		

Note: Regression yielding the mean number of marbles in women's left hand in the treated group minus the mean number of marbles in women's left hand in the control group. Controls: strata fixed-effect, age, some formal education, muslim, mossi, baseline ipv.

Table A-3.10: Estimated prevalence with LE and direct questions by Mossi

	List Experiment	Direct	Diff	p-value	N
Panel A: Mossi					
less severe	20.99	14.3	6.7	.2494	1633
se	.059	.03			
severe	21.05	11.89	9.16	.136	1633
se	.055	.028			
rape	20.61	9.31	11.3	.0197	1633
se	.051	.027			
Panel B: not Mossi					
less severe	20.7	21.14	-.44	.9312	2867
se	.045	.022			
severe	18.35	18.75	-.4	.9369	2867
se	.047	.022			
rape	20.68	15.54	5.15	.2901	2867
se	.042	.019			

Note: Controls: strata fixed-effect, age, some formal education, Muslim. Standard errors clustered at the strata level.

Table A-3.11: Estimated prevalence with LE and direct questions by religion

	List Experiment	Direct	Diff	p-value	N
Panel A: Muslim					
less severe	22.82	14.87	7.95	.1431	2161
se	.055	.028			
severe	19.97	12.66	7.31	.1875	2161
se	.049	.024			
rape	23.17	8.64	14.53	.001	2161
se	.047	.024			
Panel B: not Muslim					
less severe	15.82	21	-5.18	.3573	2339
se	.052	.023			
severe	24.18	18.22	5.95	.3402	2339
se	.057	.029			
rape	17.66	17.65	0	.9994	2339
se	.054	.021			

Note: Controls: strata fixed-effect, age, some formal education, Mossi. Standard errors clustered at the strata level. pvalue obtained using the suest command.

Table A-3.12: Estimated prevalence with LE and direct questions by schooling

	List Experiment	Direct	Diff	p-value	N
Panel A: no formal school					
less severe	20.89	17.37	3.52	.4578	3812
se	.047	.021			
severe	17.84	13.35	4.49	.3337	3812
se	.041	.017			
rape	22.65	13.18	9.47	.0232	3812
se	.038	.021			
Panel B: some formal school					
less severe	20.17	15.45	4.73	.651	688
se	.082	.054			
severe	19.37	19.74	-.37	.9765	688
se	.116	.061			
rape	1.18	5.91	-4.72	.6927	688
se	.12	.018			

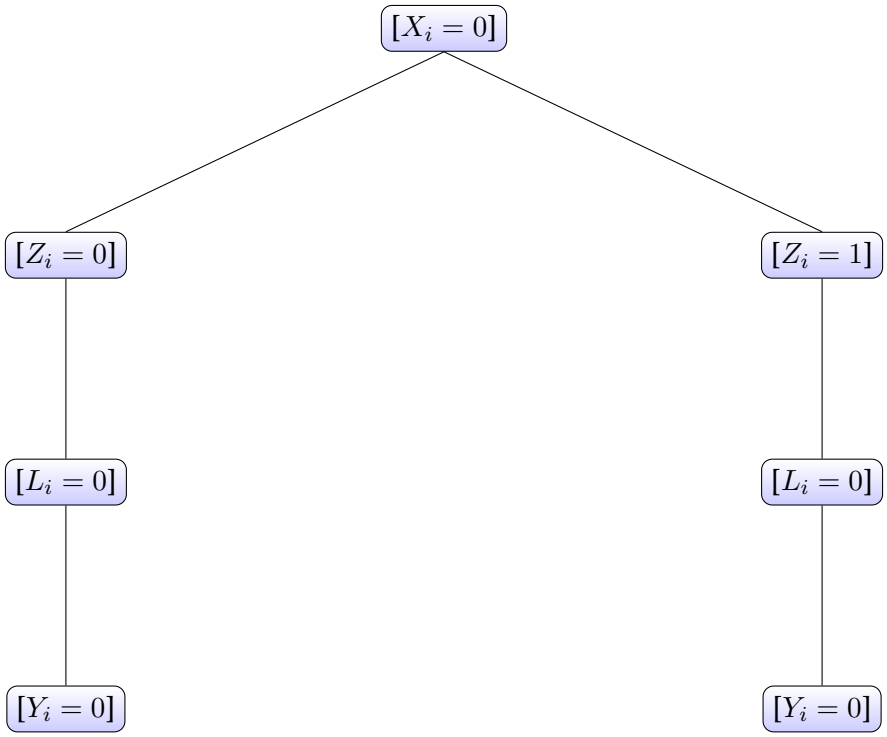
Note: Controls: strata fixed-effect, age, some formal education, Muslim, Mossi. Standard errors clustered at the strata level. pvalue obtained using the suest command in STATA.

Table A-3.13: Heterogeneous response with the LE by sex of first child

	less severe	severe	rape
Panel A: direct-reveal whose first child is a girl			
Treated with list			
	0.181	0.536	0.741**
	(0.57)	(0.36)	(0.36)
Mean Dep. Var.	1.42	0.97	1.33
pval H0=1	0.15	0.20	0.48
N	442	329	378
r2	0.83	0.87	0.81
Panel B: direct-reveal whose first child is a boy			
Treated with list			
	0.844***	0.655**	0.636**
	(0.21)	(0.27)	(0.25)
Mean Dep. Var.	1.36	0.99	1.48
pval H0=1	0.46	0.20	0.14
N	453	321	382
r2	0.72	0.79	0.85

Note: The dependent variables are the number of items women agreed with in the LE. I include strata fixed-effect and control for women's age category, whether she belongs to main ethnic group (mossi) and whether she belongs to the main religious group (muslim). Standard errors are in parenthesis.

Figure A-3.2: Framework - Women who did not experience IPV



Note:

$X_i = \begin{cases} 1 & \text{if individual i has experienced IPV} \\ 0 & \text{otherwise} \end{cases}$

$Z_i = \begin{cases} 1 & \text{if individual i randomly assigned to the baseline list + the sensitive item} \\ 0 & \text{if individual i randomly assigned to the baseline list} \end{cases}$

$L_i = \begin{cases} 1 & \text{if individual i reveals experience of IPV with the list} \\ 0 & \text{if not} \end{cases}$

$Y_i = \begin{cases} 1 & \text{if individual reveals she has experienced IPV with the direct question.} \\ 0 & \text{if not} \end{cases}$

Empirically, L_i can only be known in expectation. I suppose that women who did not experience IPV will not provide a false statement about their experience, or lack thereof, of IPV. Women who are not exposed to the baseline list + the sensitive item cannot declare with the list they were the victim of IPV.

Table A-3.14: Correlates of IPV - severe IPV

VARIABLES	(1) LE	(2) direct	(3) pval of diff.	(4) MHT adjusted p
treated with list	0.177**			
	(0.0887)			
campaign cluster	0.0265	0.00173	.76969088	1
	(0.0570)	(0.0183)		
ref cat age 18-25				
age 25-35	-0.0805	0.0270	.45358801	1
	(0.0807)	(0.0256)		
age 35-45	-0.0170	-0.00734	.9484332	1
	(0.0862)	(0.0275)		
age 45+	0.0182	0.0325	.93451501	1
	(0.0997)	(0.0313)		
some formal school	0.103	0.0672***	.76167476	1
	(0.0695)	(0.0222)		
girl first child	0.0649	0.0334**	.75206033	1
	(0.0517)	(0.0167)		
muslim	-0.0342	-0.0274	.93329526	1
	(0.0614)	(0.0198)		
mossi	0.0190	-0.0613***	.31801757	1
	(0.0598)	(0.0194)		
polygamous	0.0160	0.0138	.98112922	1
	(0.0559)	(0.0181)		
Observations	4,384	1,751		
R-squared	0.039	0.019		
mean dep var	1.042	0.138		
pval Joint significance test	0.6833	0.0001	.96625089	

Note In Column 1, the dependent variable is the number of items agreed upon by the respondent. Column 1 displays the coefficients on the interaction terms of equation 3.4 ie ρ_j. In Column 2, the dependent variable is the binary response to whether women experienced IPV when asked directly. Column 2 displays the coefficients δ of equation 3.5 for the control group only. Column 3 provides the pvalue of the difference between ρ_j and δ yielded by the suest command of STATA and column 4 shows the same value adjsted for multiple hypothesis testing using the Bonferroni method. The line "pval joint significance test" shows the pvalue of a joint significance test of the terms of interactions in column 1, the pvalue of the F-test in column 2 and the pvalue of the joint significance of the difference between ρ_j and δ in Column 3. Standard errors clustered at the strata level are in parenthesis.

Appendix B

B-3.1. Investigating the relationship between the radio treatment and marital rape

Even though the difference in the way the radio treatment correlates under the LE and under the direct measure of marital rape is not statistically significant once I correct for MHT, column 1 of table 3.12 nevertheless shows that IPV prevalence is higher among women residing in campaign clusters compared to women residing in noncampaign clusters. In order to refine the understanding of the relationship between the radio treatment and marital rape, I exploit an additional source of variation: the individual distribution of radio in both campaign and noncampaign clusters. In Pouliquen *et al.* (2021), between March and June 2017, eligible women were randomly distributed radios through a lottery. As a result, women can be broken down in 3 categories: women who are treated individually by receiving a radio at random ($N = 991$ in my sample), women who are not treated individually ($N = 1000$), and women who were not eligible to participate in the lottery ($N = 2721$). Women were ineligible for the lottery if they already owned or had access to a radio at baseline, which is probably not random. Owning or having access to a radio at baseline can yet be used as a proxy of the length of exposure to the radio treatment at the cluster level.

Table B-3.1 suggests that there is a differential effect of the individual radio treatment on women's declaration of marital rape. Even though it is only marginally significant, women exposed to the radio treatment and residing in the campaign clusters are less likely to declare they experienced marital rape than women ineligible to the radio treatment in campaign clusters. The prevalence of marital rape among women who are not eligible to the individual radio treatment and residing in campaign clusters is 33%, whereas the prevalence among women who were exposed to the individual treatment in campaign clusters is 16% (the difference between the two is statistically significant ($pval = 0.0827$)). As women were ineligible because they already owned a radio before the lottery, it could suggest that the longer the exposure to the radio treatment the higher the declaration of marital rape with the LE.

Pouliquen *et al.* (2021) find that women in noncampaign areas who received a radio reduced contraception use by 5.2 percentage points and had more conservative gender attitudes, and that, in contrast, modern contraceptive uptake rose by 5.9 percentage points in campaign areas and went up by nearly as much among women who received a radio in campaign areas.

Table B-3.1: Individual treatment, campaign clusters and marital rape

VARIABLES	Marital rape
treated with list	0.139***
	(0.0523)
campaign cluster	-0.00531
	(0.0687)
treated with list x campaign cluster	0.197**
	(0.0846)
Individual treatment ref: Non-eligible (already has a radio)	
no individual radio treatment	-0.0766
	(0.0867)
individual radio treatment	-0.0551
	(0.0952)
treated with list x no individual radio treatment	0.114
	(0.115)
treated with list x individual radio treatment	-0.00846
	(0.0989)
campaign cluster x no individual radio treatment	-0.108
	(0.123)
campaign cluster x individual radio treatment	0.111
	(0.124)
treated with list x campaign cluster x no indiv. radio treatment	-0.200
	(0.173)
treated with list x campaign cluster x indiv. radio treatment	-0.225‡
	(0.150)
Observations	4,384
Controls	YES
R-squared	0.035
pval test: LE declaration for indiv. treated in campaign clusters vs ineligible in campaign clusters	0.0827
pval test: LE declaration noneligible in campaign clusters vs noneligible in noncampaign clusters	0.0153
pval test: LE declaration for indiv. treated in campaign clusters = 0	0.0878

Note: The dependent variable is the number of items women agreed with in the LE. I include controls for women's age category, whether she belongs to main ethnic group (mossi) and whether she belongs to the main religious group (muslim). Standard errors are in parenthesis and clustered at the strata level. Significance level: $***p < 0.01, **p < 0.05, *p < 0.1, ‡p < 0.14$

However, they do not find any effect of the treatment on family planning variables such as the number of desired children or birth spacing.

With the results of Pouliquen *et al.* (2021) in mind, I suggest three potential pathways that could explain the relationship between IPV measured under the list and the radio treatment. First, the radio treatment may increase conflict around contraception use with the partner, which would result in more experience of IPV. Let's call this pathway the experience channel. Second, because the radio treatment is rather comprehensive and touches upon topics such as mutual respect and joint decision-making with the partner, it may improve women's awareness of unacceptable behaviors from their partner, the radio treatment may increase their ability to identify marital rape. Let's call this pathway the identification channel. Finally, elements in the radio treatment may make women feel more empowered and increase their willingness to reveal IPV. Let's call this pathway the report channel. Those pathways are not mutually exclusive.

In order to investigate the report channel, I use the test inspired by Aronow *et al.* (2015), as in section 3.6.5.

Panels A and B of table B-3.2 describes the extent to which women who revealed IPV with the direct question also revealed IPV with the LE for the three types of IPV. Overall, in campaign clusters, direct-reveal women appear to have complied with the protocol of the LE and declared IPV with the list experiment. For instance, nearly 90% of women who declared they experienced marital rape with the direct question used the LE to declare this type of IPV as well. By contrast, Panel B of table B-3.2 may suggest that for marital rape, only 35% of women residing in noncampaign clusters comply with the protocol of the LE, as there is a dip in the share of direct-reveal women who used the list to declare marital rape. The difference in compliance between campaign clusters and noncampaign clusters is statistically significant. As a result, it seems that some women respond to the radio treatment through the report channel: the radio treatment combined with the plausible deniability of the LE may encourage some women to report more marital rape but the LE on its own, is not enough to encourage all women to declare the first time they are asked about what probably is the most sensitive form of IPV measured here. This result is also consistent with the results of Pouliquen *et al.* (2021): they find that women in noncampaign areas exhibit more conservative gender attitudes, which may be a signal of a lesser willingness to reveal IPV.

Isolating the impact of the identification channel from the report channel is challenging because even though women are more likely to identify their experienced marital tape, if their willingness to reveal it remains unchanged, it will not show in the estimates. Shreds of evidence could suggest that the report channel might be the main channel driving women's increased declaration in marital rape rather than the identification channel. Indeed, if the increased report in marital rape experience was at least partly driven by the identification channel, I could expect to see an increase in the probability to reveal marital rape under the direct question among women who were exposed to the radio treatment. Even though it is only a weak signal of it, the second column of table 3.12 shows that women exposed to the radio treatment are not more likely to declare they experienced marital rape under the direct measure.

Finally, I hypothetized that the radio treatment may have an impact on IPV through a disagreement about the use of contraception between partners. In order to test this channel, I compare the prevalence of IPV under the list according to whether women use implants and injectables, which are the most used method of contraception in the sample, at the time of the survey. The use of contraceptives may be endogenous but if the experience channel does not matter here, I would not expect a differential impact of contraception use on IPV among women residing in

Table B-3.2: Compliance to the LE protocol by cluster status

	less severe	severe	rape
Panel A: direct-reveal in treated clusters			
Treated with list			
	0.685*	0.626***	0.894***
	(0.36)	(0.19)	(0.29)
Mean Dep. Var.	1.29	1.05	1.26
pval H0=1	0.38	0.05	0.72
N	438	333	323
r2	0.70	0.76	0.78
Panel B: direct-reveal in non-treated clusters			
Treated with list			
	0.639***	0.567***	0.349
	(0.18)	(0.20)	(0.22)
Mean Dep. Var.	1.51	0.89	1.53
pval H0=1	0.05	0.03	0.00
pval diff. in compliance with treated clusters	0.81	0.97	0.07
N	477	332	458
r2	0.61	0.65	0.51

Note: The dependent variables are the number of items women agreed with in the LE. I include controls for women's age category, whether she belongs to main ethnic group (mossi) and whether she belongs to the main religious group (muslim). Standard errors are in parenthesis and clustered at the strata level. Significance level: $***p < 0.01, **p < 0.05, *p < 0.1$.

campaign clusters. Panel A of table B-3.3, however, suggests that there is a positive association between experiencing marital rape and using modern contraception. The prevalence of marital rape among women residing in campaign clusters and who use modern contraception is 36%, whereas it is 20% for women residing in campaign clusters, and who do not use modern contraception (the pvalue of the diffence is 0.052).

B-3.2. Weather variation and measures of IPV

In this section, I explore whether weather shocks correlate differently with IPV according to the method used. The exercise is done separately from the rest of the analysis because some of the respondents in the original sample were not geolocalized.

The literature exploring the link between weather shocks and IPV in sub-Saharan Africa finds no robust evidence that droughts increase IPV (Cools *et al.* (2020)). The ability to capture an impact of weather shock may nevetheless depend on the way IPV is measured. To measure weather shocks, I use the the standardized precipitation evapotranspiration index (SPEI) developed by Vicente-Serrano *et al.* (2010). The SPEI index is a climatic drought index that takes into account precipitations and temperatures.[21] Weather data are recorded monthly on a 0.5×0.5

[21] Unlike the standardized precipitation index (SPI) that only takes into account precipitation.

Table B-3.3: Correlation between contraception use and IPV under the LE

VARIABLES	(1) rape
treated with list	0.184***
	(0.0523)
contraception	0.0809
	(0.0704)
campaign clusters	0.0749
	(0.0751)
treated with list x contraception	-0.0846
	(0.0944)
treated with list x campaign clusters	-0.0478
	(0.0891)
contraception x campaign clusters	-0.238*
	(0.139)
treated with list x campaign clusters x contraception	0.393**
	(0.169)
Observations	4,384
R-squared	0.029
mean dep var	1.212
pval test: prevalence in campaign clusters with contraception = prevalence in campaign clusters no contraception	0.052
pval test: prevalence in noncampaign clusters with contraception = prevalence in noncampaign clusters no contraception	0.955

Note: The dependent variables are the number of items women agreed with in the LE. I include controls for women's age category, whether she belongs to main ethnic group (mossi) and whether she belongs to the main religious group (muslim). Standard errors are in parenthesis and clustered at the strata level. Significance level: $***p < 0.01, **p < 0.05, *p < 0.1$.

degree (0.5 degree corresponds to approximately 56 kilometers at the Equator). SPEI values range from -3 to 3, and the categorization of dryness/wetness is as follows: extreme wetness (more than 2), very wet (1.50 to 1.99), moderately wet (1 to 1.49) near normal (-1 to 1), moderate dryness (-1.49 to -1), severe dryness (-1.99 to -1.5), and extreme dryness (less than -2).

I use the SPEI4, which measures the anomaly in the cumulated precipitation and temperature over four months. For instance, the SPEI4 for September 2018 measures the anomaly in precipitation and temperature in June, July, August and September, compared to the average level of precipitation and temperature over 30 years. The SPEI4 for September covers the main rainy season in Burkina Faso, that lasts from June to September.[22] Rainfall in those months are crucial to ensure a good harvest as crops are mostly rainfed. A positive weather shock is very likely to translate into an increase in income for rural households.

I study the relationship between weather variations and IPV shortly after the 2018 rainy season (follow-up data was collected in November 2018). Values of the SPEI4 for September 2018 range from 0.12 to 1.65, which suggests that the 2018 rainy season was considered normal to very wet. As shown in the fifth column of table B-3.4, I find that a 1 standard deviation increase

[22]May and October are months of transition.

in the SPEI index is associated with less report of marital rape (the coefficient on the interaction term is -0.2 and significant at the 10% level).[23] The comparison between the interaction term in column 5 and the coefficient in column 6 of table B-3.4 suggests that the LE measure allows to capture an association between weather variation and marital rape, whereas the direct measure does not. I find no relation between weather variation and less severe or severe physical violence, whether I use the direct measure of IPV, or the measure yielded by the list experiment.

Table B-3.4: Weather and IPV

VARIABLES	less severe		severe		marital rape	
	LE	direct	LE	direct	LE	direct
treated with list	0.0706		0.205		0.417***	
	(0.139)		(0.139)		(0.114)	
SPEI September 2018	0.0889	0.0740	0.137	-0.0348	0.124	0.0240
	(0.0915)	(0.0485)	(0.102)	(0.0468)	(0.0829)	(0.0518)
treated with list x SPEI Sept 2018	0.125		-0.0103		-0.202*	
	(0.131)		(0.126)		(0.109)	
Observations	3,809	1,461	3,809	1,461	3,809	1,461
mean SPEI4 2018	1.06	1.06	1.06	1.06	1.06	1.06
mean dep var	1.342	1.342	1.042	1.042	1.261	1.261
pvalue of difference btw LE and direct		0.72		0.85		0.06

Note: In columns 1, 3, and 5, the dependent variable is the number of items agreed upon by the respondent. . In Column 2, 4, and 6 the dependent variable is the binary response to whether women experienced IPV when asked directly. Controls: residing in a campaign cluster, age, some formal education, muslim, mossi, and whether the partner is polygamous. Significance level: $***p < 0.01, **p < 0.05, *p < 0.1$.

[23] The SPEI index values are standardized over the 0.5×0.5 degree area, over which precipitation and temperature are measured. Because I retrieve the value of the SPEI4 for the precise geolocation of respondents, the mean of the SPEI4 in my sample is different than zero, and the standard deviation is different than 1.

Appendix C

C-3.1. Heterogeneity in the treatment effect on the measurement error

Again, I am interested in the impact of X on Y^*, so I run the following linear regression: $Y^* = \beta * X + u$. I assume that X is exogenous to Y^* so that $cov(X, u) = 0$ and $E(u|X) = 0$. I cannot perfectly observe the experience of IPV of women Y^*, but I observe the report of the experience of IPV: Y, with a level of error v, so that:

$$Y = Y^* + v \tag{3.6}$$

I assume that the measurement error v consists in two components: a random component, v', and a component that is correlated with Y^* and with the variable X is a nonlinear way, so that:

$$v = \rho * Y^* + 1[X > \mu](\delta_1 X + v') + 1[X \leq \mu](\delta_2 X + v') \tag{3.7}$$

With

$$1[X > \mu] = \begin{cases} 1 & if X > \mu \\ 0 & otherwise \end{cases}$$

To ease the reading, I note: $1[X > \mu] = \Pi$.

In practice, I run the following regression: $Y = \beta_b * X + u$. By definition:

$$\beta_b = \frac{cov(Y, X)}{V(X)}$$

Substituting Y for Y^* and v:

$$\beta_b = \frac{cov(\beta * X + u + \rho * Y^* + \Pi * (\delta_1 * X + v') + (1 - \Pi)(\delta_2 * X + v'), X)}{V(X)}$$

$$\iff \beta_b = \beta + \frac{cov(u,X)}{V(X)} + \rho * \frac{cov(Y^*,X)}{V(X)} + \frac{cov(\Pi*(\delta_1*X+v'),X)}{V(X)} + \frac{(1-\Pi)(\delta_2*X+v'),X)}{V(X)}$$

$$\iff \beta_b - \beta = \rho * \beta + \delta_1 * \frac{cov(\Pi*X,X)}{V(X)} + \delta_2 * \frac{cov((1-\Pi)*X,X)}{V(X)}$$

Using the formula of the covariance:

$cov(\Pi * X, X) = E(\Pi * X^2) - E(\Pi * X)E(X)$

It stems from the definition of Π that:

$$\Pi * X^2 = \begin{cases} 0 & if \Pi = 0 \\ X^2 & if \Pi = 1 \end{cases}$$

Decomposing each term of the covariance formula yields:

$E(\Pi * X^2) = E(X^2|\Pi = 1)P(\Pi = 1)$

$E(\Pi * X) = E(X|\Pi = 1)P(\Pi = 1)$

It results from the above that:

$cov(\Pi * X, X) = E(X^2|\Pi = 1)P(\Pi = 1) - E(X|\Pi = 1)P(\Pi = 1)E(X)$

Using the law of total expectation:

$$E(X) = E(X|\Pi = 1)P(\Pi = 1) + E(X|\Pi = 0)P(\Pi = 0)$$

Reinjecting the expressions above:

$cov(\Pi * X, X) = E(X^2|\Pi = 1)P(\Pi = 1) - E(X|\Pi = 1)P(\Pi = 1)[E(X|\Pi = 1)P(\Pi = 1) + E(X|\Pi = 0)P(\Pi = 0)]$

$\iff cov(\Pi * X, X) = P(\Pi = 1)[E(X^2|\Pi = 1) - E(X|\Pi = 1)^2] + E(X|\Pi = 0)P(\Pi = 0)$

$\iff cov(\Pi * X, X) = P(\Pi = 1)V(X|\Pi = 1) + E(X|\Pi = 0)P(\Pi = 0)$

As a result:

$\beta_b - \beta = \rho * \beta + \delta_1 * E(\Pi)\frac{V(X|\Pi=1)+E(X|\Pi=0)P(\Pi=0)}{V(X)} + \delta_2 * \frac{cov((1-\Pi)*X,X)}{V(X)}$

Using again the formula of the covariance:

$$cov((1 - \Pi * X), X) = E((1 - \Pi)X^2) - E((1 - \Pi)X)E(X)$$

It stems from the definition of $1 - \Pi$ that:

$$(1 - \Pi)X^2 = \begin{cases} 0 & if 1 - \Pi = 0 \\ X^2 & if 1 - \Pi = 1 \end{cases}$$

It follows that:

$E((1 - \Pi)X^2) = E(X^2|\Pi = 0)P(\Pi = 0)$

$E((1 - \Pi)X)E(X) = (E(X|\Pi = 0))^2 P(\Pi = 0)$

And that:

$cov((1 - \Pi * X), X) = P(\Pi = 0)V(X|\Pi = 0) + E(X|\Pi = 1)P(\Pi = 1)$

Finally:

$$\begin{aligned} \beta_b - \beta = \rho * \beta + \delta_1 * E(\Pi)\frac{V(X|\Pi = 1) + E(X|\Pi = 0)P(\Pi = 0)}{V(X)} \\ +\delta_2 * \frac{P(\Pi = 0)V(X|\Pi = 0) + E(X|\Pi = 1)P(\Pi = 1)}{V(X)} \end{aligned} \tag{3.8}$$

CONCLUSION

General Conclusion

Main results

The dissertation provides new evidence that advance the understanding of the sources of fragility and strength of women in the family, in part proxied by women's tolerance and experience of domestic violence. Chapter 1 shows that having more primary education improves the condition in which women join the family of their partner: it delays the age at marriage, the age at first child, the tolerance of IPV, and the exposure to IPV. The change is driven by women's own education, and not by the education of their partner. In chapter 2, I look at whether, once they have joined the household of their partner, having sons strenghtens the status of their mother. The key mechanism is a preference for sons, who are the pillar of the lineage in patrilineal societies (Lesthaeghe (1989), Milazzo (2014)). Even though I am unable to provide a clear-cut answer to the question, I find descriptive evidence that the impact of having sons may differ according to the type of union monogamous or polygamous , and the rank of the spouse among women married to a polygamous partner; it may improve polygamous women's outcomes but may worsen monogamous women's outcomes. Chapter 3 critically discusses the impact of using only direct measures of IPV to measure domestic violence. It shows that the direct measures of IPV underestimate the prevalence of the most intense forms of violence by 7 to 9 percentage points, and that it modifies the diagnostic of the characteristics that seem to make women vulnerable to domestic violence.

Policy Implications

The dissertation has several policy implications. Chapter 1 shows that, even policies that did not directly target women's marital outcomes may have beneficial impacts beyond the initial goals of the policy. It also shows that even in contexts where the quality of education tends to be low, attending primary education is nevertheless enough to shield women from experi-

encing IPV later in their life cycle. It further calls for an intensification of the efforts already engaged to reach universal education, in particular after the covid crisis, which is feared to cause an increase in drop-out rates among girls of schooling age.

Chapter 2 suggests that, even in a patrilineal society like Burkina Faso, the role of sons ont their mother's status, autonomy, and access to resources might be ambiguous and complex. The results of chapter 2 also invites to question the consistency of the measures used in large household surveys to measure women's status in the household, the part they take in decision-making, and how it may relate to their actual empowerment. The question of the importance of measurement is also crucial for the experience of IPV, and the discussions of chapter 2 have to be put in balance with the results of chapter 3, which shows that, when using a tool meant to get closer to a more accurate prevalence of IPV like a list experiment, being the mother of a first born son does seem to protect women against experiencing IPV. It is important to note that in Burkina Faso in 2018, women had on average 5.19 children,[24] which makes the country a high fertility context. As decreasing fertility has been an objective of public policies in sub-Saharan Africa for several years,[25] it is an open-question how the impact of the sex composition of children may become geared towards son preference as women will tend to have less children. To anticipate this risk, policy makers could act to further alleviate the concerns that may drive women to rely on boys to be insured against adverse shocks by implementing safety nets to decouple women's economic security from their descendants.

Finally, chapter 3 invites policy makers to recognize that domestic violence is not sporadic nor rare. Domestic violence is an egregious form of the denial of women's rights and it is frequent enough to justify investments to fight against it. Though the following suggestions may require further research to refine our understanding of their efficiency, several non-mutually exclusive paths could be explored. First, in countries where it is needed, reform family laws to improve women's access to property rights, and to inheritance rights, in order to secure an autonomous and sustainable access to resources for women. In countries where the law have already done that, encourage women to marry under civil law in addition to customary laws, in order to strenghen their claim (and hence their threat point) in case widowhood or divorce, especially from a violent partner. Funding women shelters could also be a way to provide women with a quick and temporary way out in case their partner is violent. Second, policy makers could greatly improve the way the judicial and the health system treat women who experience IPV. From the qualitative work I led in Burkina Faso and based on exchanges with the

[24] Source: World Bank databank.

[25] In 2013, 83% of African countries had targets to decrese fertilityNations (2013).

Association of the female legal experts of Burkina Faso, domestic violence is very often seen as private matter that should be taken care of within the family by adhoc interventions. According to testimonies, the quality of the reception of victims of IPV by police forces is poor not to say inadequate. Inspirations could be drawn by the implementation of Domestic Violence and Victim Support Unit of the Ghana Police Service, that exclusively female units within police precincts of Accra, Ghana, and implemented to accompany victims of IPV. Finally, domestic violence persists alongside restrictive gender norms that enable violence perpetrators. Fighting against these norms through sensitization, by either targetting police forces, judges, or health professionals, or by implementing campaigns meant to reach a larger audience, through edutainment programs for instance.

Perspectives for future research

The dissertation is a stepping stone to investigate further the determinants of women's status in the family in sub-Saharan Africa.

The results of chapter 2 seem to suggest that the impact of attending primary school does not go through what women learnt at school. One hypothesis would be that the decrease in tolerance of IPV may come from evolving in a co-ed environment. If this hypothesis were true, it would suggest that the mere repeated contact with the other sex may be enough to change gender norms. The historical hindsight is not sufficient to use the data of Benin to test this hypothesis but there are countries in West Africa tat have expanded primary school before Benin did, like Nigeria for instance, that did it in 1970's. I will explore this research avenue with Rozenn Hotte (University of Tours). When the next DHS for Benin is available, we will also take the opportunity to explore the long-term effects of the contruction of primary schools.

Chapter 2 and 3 lay bare the difficulty to measure acurately women's agency and the experience of IPV. During a postdoc at Northwestern University, I will study with Seema Jayachandran (Northwestern University) on an indicator of women's agency she developed in India (Jayachandran *et al.* (2021)). To build the indicator, they first used qualitative coding methods to score each woman's agency based on the semi-directed interviews, which they treat as the true agency of women. To identify the close-ended questions most predictive of the "truth," they apply statistical algorithms and machine learning method to select five questions that are the most strongly correlated with the coded qualitative interviews. I will collaborate on a replication of the study in one or more other contexts with the goals of testing how well the method works to create accurate measure of women's agency for those contexts, understand the over-

lap in the best questions selected in different contexts, and create an optimized "universal" module that could be applied across contexts. As the method was designed to be applied to other latent constructs, applying to the experience of IPV could be an interesting research avenues, especially to inform targetting of policies aimed at fighting against IPV.

As mentionned in the policy implications of the dissertation, eroding the formation of gender norms that enable violence against women is probably privotal in the fight for gender equality. In a project joint with Yasmine Bekkouche, Rozenn Hotte (University of Tours), and Nina Buchmann (Stanford University), we will evaluate a program whose ambition is to promote gender equality in Benin by fighting gender restricitve norms. The Champion of Change (CoC) program, developed by the NGO Plan International, is an empowerment strategy that provides gender training for adolescent girls and boys in selected communities by selecting adolescents and youth to become "Champions of Change" after being trained on a series of modules, so that they can organize outreach community activities to influence their peers or members of their community to organize gender programming in treatment communities. The proposed research project intends to shed light on the optimal targeting of interventions designed to change gender norms, in particular, whether it is more effective to target and reinforce the views of adolescents who already hold gender progressive gender views, or to target and challenge the views of adolescents who hold less progressive gender views.

I consider the dissertation as a first step in a more general research project aiming at understanding the decision process in non-nuclear families. In that vein, I will explore further this question during my postdoc by working on the imapcts pf a cash transfer program in Mali. Chris Udry (Northwestern University) and Lori Beaman (Northwestern University) have collected data from an randomized controlled trial (RCT) in Mali, which provided unconditional cash transfers to either an adult female or her husband. They find that the identity of the recipient reshapes production activities on the plots controlled by women and men in the household, shifts expenditure across different goods, and causes changes in the allocation in productive resources between the nuclear and extended families. I will work with them to couple the experimental variation with natural variation in growing conditions to provide another dimension of within family reallocations to permit the testing of generalized collective models of the household.

Following a close research agenda, I will join Kenneth Houngbedji (IRD, DIAL), and Liam Wren-Lewis (PSE) to work on a project on land formalization. Between 2009 and 2011, the gov-

ernment of Benin formalized rural land rights in 300 randomly selected villages. Evidence from survey data conducted in a subsample of treatment and control villages suggests this helped to empower women: female-managed plots were more likely to be left fallow (Goldstein *et al.* (2018)), widows were less likely to leave the village, and were more likely to report that they were involved in household decisions around land. At least two important questions remain, however. First, why did this change occur, when women were only rarely named on the formal documents? Second, did this change persist after the last round of survey data collected in 2015, or did the impact disappear once government attention on the program waned? To answer these questions, we plan to collect administrative data from tribunal and conciliation courts responsible of arbitrating land disputes in treatment and control villages.

BIBLIOGRAPHY

Agüero, J. M. and **Frisancho, V.**, 2017. "Systematic Bias in Sensitive Health Behaviors and Its Impact on Treatment Effects: An Application to Violence against Women." Tech. Rep., Inter-American Development Bank.

Aizer, A., 2010. "The Gender Wage Gap and Domestic Violence." *The American Economic Review* 100(4), 1847.

Akerlof, G. A. and **Kranton, R. E.**, 2000. "Economics and identity." *The Quarterly journal of Economics* 115(3), 715–753.

Alesina, A., **Brioschi, B.**, and **Ferrara, E. L.**, 2016. "Violence against women: A cross-cultural analysis for Africa." Tech. Rep., National Bureau of Economic Research.

Anderson, M. L., 2008. "Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American statistical Association* 103(484), 1481–1495.

Anderson, S. and **Genicot, G.**, 2015. "Suicide and Property Rights in India." *Journal of Development Economics* 114, 64–78.

Andrew, 2014. "Thinking of doing a list experiment? Here's a list of reasons why you should think again." Tech. Rep., https://statmodeling.stat.columbia.edu/2014/04/23/thinking-list-experiment-heres-list-reasons-think/.

Annan, J., **Donald, A.**, **Goldstein, M.**, **Martinez, P. G.**, and **Koolwal, G.**, 2021. "Taking Power: Women's Empowerment and Household Well-being in Sub-Saharan Africa." *World Development* 140, 105292.

Aronow, P. M., **Coppock, A.**, **Crawford, F. W.**, and **Green, D. P.**, 2015. "Combining list experiment and direct question estimates of sensitive behavior prevalence." *Journal of Survey Statistics and Methodology* 3(1), 43–66.

Baird, S., **McIntosh, C.**, and **Özler, B.**, 2011. "Cash or condition? Evidence from a cash transfer experiment." *The Quarterly Journal of Economics* 126(4), 1709–1753.

Baland, J.-M. and **Ziparo, R.**, 2018. "Intra-household bargaining in poor countries." In "Towards Gender Equity in Development," , edited by **Anderson, S.**, **Beaman, L.**, and **Platteau, J.-P.**, chap. 4. Oxford University Press, pages 69–96.

Basu, D. and **De Jong, R.**, 2010. "Son targeting fertility behavior: Some consequences and determinants." *Demography* 47(2), 521–536.

Bellemare, M. F., **Novak, L.**, and **Steinmetz, T. L.**, 2015. "All in the family: Explaining the persistence of female genital cutting in West Africa." *Journal of Development Economics* 116, 252–265.

Bénabou, R. and **Tirole, J.**, 2006. "Incentives and prosocial behavior." *American Economic Review* 96(5), 1652–1678.

Benabou, R. and **Tirole, J.**, 2011. "Laws and norms." Tech. Rep., National Bureau of Economic Research.

Blair, G., **Chou, W.**, and **Imai, K.**, 2018. "List experiments with measurement error." Tech. Rep., Technical report. Princeton University. https://imai. princeton. edu â.

Blair, G. and **Imai, K.**, 2012. "Statistical analysis of list experiments." *Political Analysis* 20(1), 47–77.

Bledsoe, C., **Banja, F.**, and **Hill, A. G.**, 1998. "Reproductive mishaps and western contraception: an African challenge to fertility theory." *Population and Development Review* , 15–57.

Bloch, F. and **Rao, V.**, 2002. "Terror as a bargaining instrument: A case study of dowry violence in rural India." *The American Economic Review* 92(4), 1029–1043.

Bobonis, G. J., **González-Brenes, M.**, **Castro, R.**, *et al.*, 2013. "Public transfers and domestic violence: The roles of private information and spousal control." *American Economic journal: Economic Policy* 5(1), 179–205.

Boltz, M. and **Chort, I.**, 2019. "The risk of polygamy and wives' saving behavior." *The World Bank Economic Review* 33(1), 209–230.

Boyle, M. H., **Georgiades, K.**, **Cullen, J.**, and **Racine, Y.**, 2009. "Community influences on intimate partner violence in India: Women's education, attitudes towards mistreatment and standards of living." *Social Science & Medicine* 69(5), 691–697.

Breierova, L. and **Duflo, E.**, 2004. "The Impact of Education on Fertility and Child Mortality: Do Fathers Really Matter Less Than Mothers?" *NBER Working Paper* (10513).

Browning, M., **Bourguignon, F.**, **Chiappori, P.-A.**, and **Lechene, V.**, 1994. "Income and Outcomes: A Structural Model of Intrahousehold Allocation." *journal of Political Economy* 102(6), 1067–1096.

Bulte, E. and **Lensink, R.**, 2019. "Women's empowerment and domestic abuse: Experimental evidence from Vietnam." *European economic review* 115, 172–191.

Bursztyn, L., **Fujiwara, T.**, and **Pallais, A.**, 2017. "'Acting Wife': Marriage Market Incentives and Labor Market Investments." *American Economic Review* 107(11), 3288–3319.

Bursztyn, L. and **Jensen, R.**, 2017. "Social image and economic behavior in the field: Identifying, understanding, and shaping social pressure." *Annual Review of Economics* 9, 131–153.

Cameron, L. A., **Malcolm Dowling, J.**, and **Worswick, C.**, 2001. "Education and labor market participation of women in Asia: Evidence from five countries." *Economic Development and Cultural Change* 49(3), 459–477.

Cannonier, C. and **Mocan, N.**, 2018. "The Impact of Education on Women's Preferences for Gender Equality: Evidence from Sierra Leone." *Journal of Demographic Economics* 84(1), 3–40.

Card, D., **Lee, D.**, **Pei, Z.**, and **Weber, A.**, 2012. "Nonlinear policy rules and the identification and estimation of causal effects in a generalized regression kink design." Tech. Rep., National Bureau of Economic Research.

Card, D., **Lee, D. S.**, **Pei, Z.**, and **Weber, A.**, 2015. "Inference on Causal Effects in a Generalized Regression Kink Design." Tech. Rep., W.E. Upjohn Institute. DOI: 10.17848/wp15-218.

Cavalin, C., 2010. "Comment questionner les violences subies?" *Violences et santé en France* , 17.

—, 2013. "Interroger les femmes et les hommes au sujet des violences conjugales en France et aux États-Unis: entre mesures statistiques et interprétations sociologiques." *Nouvelles questions féministes* 32(1), 64–76.

Chou, W., 2018. "Lying on surveys: Methods for list experiments with direct questioning." Tech. Rep., Technical report, Princeton University.

Chou, W., **Imai, K.**, and **Rosenfeld, B.**, 2017. "Sensitive survey questions with auxiliary information." *Sociological Methods & Research* , 0049124117729711.

Chuang, E., **Dupas, P.**, **Huillery, E.**, and **Seban, J.**, 2021. "Sex, lies, and measurement: Consistency tests for indirect response survey methods." *Journal of Development Economics* 148, 102582.

Cools, S., **Flatø, M.**, and **Kotsadam, A.**, 2020. "Rainfall shocks and intimate partner violence in sub-Saharan Africa." *Journal of Peace Research* 57(3), 377–390.

Corstange, D., 2009. "Sensitive questions, truthful answers? Modeling the list experiment with LISTIT." *Political Analysis* 17(1), 45–63.

Cullen, C. A., 2020. "Method Matters: Underreporting of Intimate Partner Violence in Nigeria and Rwanda." *World Bank Policy Research Working Paper* (9274).

De Mel, S., **McKenzie, D.**, and **Woodruff, C.**, 2009. "Are Women More Credit Constrained? Experimental Evidence on Gender and Microenterprise Returns." *American Economic journal: Applied Economics* 1(3), 1–32.

Deaton, A., 1989. "Looking for boy-girl discrimination in household expenditure data." *The World Bank Economic Review* 3(1), 1–15.

Deininger, K., 2003. "Does cost of schooling affect enrollment by the poor? Universal Primary Education in Uganda." *Economics of Education Review* 22, 291–305.

DeKeseredy, W. and **Kelly, K.**, 1993. "The incidence and prevalence of woman abuse in Canadian university and college dating relationships." *Canadian journal of Sociology/Cahiers canadiens de sociologie* , 137–159.

DeKeseredy, W. S. and **Schwartz, M. D.**, 1998. "Measuring the extent of woman abuse in intimate heterosexual relationships: A critique of the Conflict Tactics Scales." *US Department of Justice Violence Against Women Grants Office Electronic Resources* .

Dercon, S. and **Krishnan, P.**, 2000. "In sickness and in health: Risk sharing within households in rural Ethiopia." *journal of Political Economy* 108(4), 688–727.

Deschênes, S., 2021. "Using List Experiments to measure Intimate Partner Violence (IPV): Lessons from rural Burkina Faso." Tech. Rep., Paris School of Economics.

Deschênes, S., **Dumas, C.**, and **Lambert, S.**, 2020. "Household resources and individual strategies." *World Development* 135, 105075.

Diallo, S. A. and **Voia, M.**, 2016. "The Threat of Domestic Violence and Women Empowerment: The Case of West Africa." *African Development Review* 28(1), 92–103.

Dobash, R. E. and **Dobash, R.**, 1979. *Violence against wives: A case against the patriarchy*. Free Press New York.

Dobash, R. E. and **Dobash, R. P.**, 1998. *Rethinking violence against women*, vol. 9. Sage Publications.

Dobash, R. P., **Dobash, R. E.**, **Wilson, M.**, and **Daly, M.**, 1992. "The myth of sexual symmetry in marital violence." *Social problems* 39(1), 71–91.

Donald, A., **Doss, C.**, **Goldstein, M. P.**, and **Gupta, S.**, 2021. "Sharing Responsibility through Joint Decision Making and Implications for Intimate-Partner Violence." .

Downey, D. B., **Jackson, P. B.**, and **Powell, B.**, 1994. "Sons versus daughters: Sex composition of children and maternal views on socialization." *The Sociological Quarterly* 35(1), 33–50.

Droitcour, J., **Caspar, R. A.**, **Hubbard, M. L.**, **Parsley, T. L.**, **Visscher, W.**, and **Ezzati, T. M.**, 2004. "The item count technique as a method of indirect questioning: A review of its development and a case study application." *Measurement errors in surveys* , 185–210.

Duflo, E., 2001. "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment." *American Economic Review* 4(91), 795–813.

—, 2003. "Grandmothers and Granddaughters: Old-Age Pensions and Intrahousehold Allocation in South Africa." *The World Bank Economic Review* 17(1), 1–25.

Duflo, E. and **Udry, C.**, 2004. "Intrahousehold Resource Allocation in Cote d'Ivoire: Social Norms, Separate Accounts and Consumption Choices." Working Paper 10498, National Bureau of Economic Research.

Duraisamy, P., **James, E.**, **Lane, J.**, and **Tan, J.-P.**, 1998. "Is there a quantity-quality trade-offs a pupil-teacher ratios increase? Evidence from Tamil Nadu, India." *International Journal of Educational Development* 18(5), 367–383.

Eady, G., 2017. "The statistical analysis of misreporting on sensitive survey questions." *Political Analysis* 25(2), 241–259.

Equipe nationale du Bénin, U. and **de Dakar, I. P.**, 2014. "Rapport d'état du Système Educatif Béninois, Pour une Revitalisation de la Politique Educative dans le Cadre du Programme Décennal de Développement du Secteur de l'Education." Tech. Rep., UNESCO.

Erten, B. and **Keskin, P.**, 2018. "For Better or for Worse?: Education and the Prevalence of Domestic Violence in Turkey." *American Economic Journal: Applied Economics* 10(1), 64–105.

Falen, D. J., 2011. *Power & Paradox: authority, Insecurity, and Creativity in Fon Gender Relations*. Africa World Press.

Flatø, M. and **Kotsadam, A.**, 2014. "Droughts and gender bias in infant mortality in Sub-Saharan Africa." Tech. Rep., Memorandum, Department of Economics, University of Oslo.

Flavin, P. and **Keane, M.**, 2009. "How angry am I? Let me count the ways: Question format bias in list experiments." *Unpublished Manuscript* , 1–17.

Folbre, N., 1986. "Cleaning house: New perspectives on Households and Economic Development." *journal of Development Economics* 22(1), 5–40.

Gastineau, B., **Gnele, J.**, and **Mizochounnou, S.**, 2015. "Pratiques scolaires et genre dans les écoles primaires à Cotonou." *Autrepart* (2), 3–22.

Gelles, R. J. and **Straus, M. A.**, 1979. "Violence in the American family." *Journal of Social Issues* 35(2), 15–39.

Ghilagaber, G. and **Elisa, W.**, 2014. "A Family of Flexible Parametric Duration Functions and Their Applications to Modeling Child-Spacing in Sub-Saharan Africa." In "Advanced Techniques for Modelling Maternal and Child Health in Africa," Springer, pages 185–209.

Glynn, A. N., 2013. "What can we learn with statistical truth serum? Design and analysis of the list experiment." *Public Opinion Quarterly* 77(S1), 159–172.

Goldstein, M., **Houngbedji, K.**, **Kondylis, F.**, **O'Sullivan, M.**, and **Selod, H.**, 2018. "Formalization without certification? Experimental evidence on property rights and investment." *Journal of Development Economics* 132, 57–74.

Gordon, M., 2000. "Definitional Issues in Violence Against Women Surveillance and Research from a Violence Research Perspective." *Violence Against Women* 6(7), 747–783.

Gray, J. P., 1998. "Ethnographic atlas codebook." *World Cultures* 10(1), 86–136.

Grémy, J.-P., 2006. "L'enquête nationale de victimation nord américaine (National Crime Survey)." *Cahiers du Centre d'Analyse et de Mathématiques sociales (Histoire du Calcul des Probabilités et de la Statistique)* 252, 70.

Hanmer, L. and **Klugman, J.**, 2016. "Exploring Women's agency and empowerment in developing countries: where do we stand?" *Feminist Economics* 22(1), 237–263.

Heath, R. and **Tan, X.**, 2018. "Worth fighting for: Daughters improve their mothers' autonomy in South Asia." *Journal of Development Economics* 135, 255–271.

Hoddinott, J. and **Haddad, L.**, 1995. "Does Female Income Share Influence Household Expenditures? Evidence from Côte d'Ivoire." *Oxford Bulletin of Economics and Statistics* 57(1), 77–96.

Hoenig, J. M. and **Heisey, D. M.**, 2001. "The abuse of power: the pervasive fallacy of power calculations for data analysis." *The American Statistician* 55(1), 19–24.

Holtzworth-Munroe, A., 2000. "A typology of men who are violent toward their female partners making sense of the heterogeneity in husband violence." *Current Directions in Psychological Science* 9(4), 140–143.

ICF, 2017. "Demographic and health survey interviewerâs manual." *MEASURE DHS* .

Imai, K., 2011. "Multivariate regression analysis for the item count technique." *journal of the American Statistical Association* 106(494), 407–416.

Iyengar, R. and **Ferrari, G.**, 2015. "Discussion sessions coupled with microfinancing may enhance the role of women in household decision-making in Burundi." In "African Successes: Human Capital, volume 2," University of Chicago Press.

Jayachandran, S., **Biradavolu, M.**, and **Cooper, J.**, 2021. "Using machine learning and qualitative interviews to design a five-question women's agency index." Tech. Rep., National Bureau of Economic Research.

Jensen, R. and **Thornton, R.**, 2003. "Early female marriage in the developing world." *Gender & Development* 11(2), 9–19.

Jewkes, R., 2002. "Intimate partner violence: causes and prevention." *The Lancet* 359(9315), 1423–1429.

Johnson, H., 1998. *Rethinking survey research on violence against women*. Sage Publications, Inc.

Johnson, M. P., 1995. "Patriarchal terrorism and common couple violence: Two forms of violence against women." *Journal of Marriage and the Family* , 283–294.

—, 2006. "Conflict and control: Gender symmetry and asymmetry in domestic violence." *Violence against women* 12(11), 1003–1018.

Johnson, M. P. and **Leone, J. M.**, 2005. "The differential effects of intimate terrorism and situational couple violence findings from the national violence against women survey." *Journal of family issues* 26(3), 322–349.

Kabeer, N., 1999. "Resources, agency, achievements: Reflections on the measurement of women's empowerment." *Development and change* 30(3), 435–464.

—, 2005. "Gender equality and women's empowerment: A critical analysis of the third millennium development goal 1." *Gender & Development* 13(1), 13–24.

Kim, J. C., **Watts, C. H.**, **Hargreaves, J. R.**, **Ndhlovu, L. X.**, **Phetla, G.**, **Morison, L. A.**, **Busza, J.**, **Porter, J. D.**, and **Pronyk, P.**, 2007. "Understanding the impact of a microfinance-based intervention on women's empowerment and the reduction of intimate partner violence in South Africa." *American Journal of Public Health* 97(10), 1794–1802.

Kishor, S. and **Subaiya, L.**, 2008. "Understanding womens empowerment: a comparative analysis of Demographic and Health Surveys (DHS) data." .

Koenig, M. A., **Lutalo, T.**, **Zhao, F.**, **Nalugoda, F.**, **Wabwire-Mangen, F.**, **Kiwanuka, N.**, **Wagman, J.**, **Serwadda, D.**, **Wawer, M.**, and **Gray, R.**, 2003. "Domestic violence in rural Uganda: evidence from a community-based study." *Bulletin of the World Health Organization* 81(1), 53–60.

Lambert, S. and **Rossi, P.**, 2016. "Sons as widowhood insurance: Evidence from Senegal." *Journal of Development Economics* 120, 113–127.

Landais, C., 2015. "Assessing the Welfare Effects of Unemployment Benefits Using the Regression Kink Design." *American Economic Journal: Economic Policy* 7(4), 243–278.

Lesthaeghe, R. J., 1989. *Reproduction and social organization in sub-Saharan Africa*, vol. 4. Univ of California Press.

Locoh, T., 1995. *Familles africaines, population et qualité de la vie*. Ceped Paris.

Lowes, S., 2016. "Kinship systems, gender norms, and household bargaining: Evidence from the matrilineal belt." Tech. Rep., Working paper.

Macmillan, R. and **Gartner, R.**, 1999. "When she brings home the bacon: Labor-force participation and the risk of spousal violence against women." *Journal of Marriage and the Family* , 947–958.

Manser, M. and **Brown, M.**, 1980. "Marriage and household decision-making: A bargaining analysis." *International Economic Review* , 31–44.

McElroy, M. B. and **Horney, M. J.**, 1981. "Nash-bargained household decisions: Toward a generalization of the theory of demand." *International Economic Review* , 333–349.

Miedema, S. S., **Haardörfer, R.**, **Girard, A. W.**, and **Yount, K. M.**, 2018. "Women's empowerment in East Africa: Development of a cross-country comparable measure." *World Development* 110, 453–464.

Milazzo, A., 2014. *Son preference, fertility and family structure: Evidence from reproductive behavior among Nigerian women*. The World Bank.

Miller, J. D., 1984. "A New Survey Technique for Studying Deviant Behavior PhD thesis The George Washington University." .

Mosedale, S., 2005. "Assessing women's empowerment: towards a conceptual framework." *Journal of international development* 17(2), 243–257.

Nations, U., 2013. "World Population Policies." Tech. Rep., United Nations Publications.

Nour, N. M., 2006. "Health consequences of child marriage in Africa." *Emerging infectious diseases* 12(11), 1644.

Novak, L., 2016. "Persistent norms and tipping points: Female genital cutting in Burkina Faso." Tech. Rep., Mimeo, University of Minnesota.

Osili, U. O. and **Long, B. T.**, 2008. "Does female schooling reduce fertility? Evidence from Nigeria." *Journal of Development Economics* (87), 57–75.

Pambè, M. W., **Gnoumou, B.**, and **Kaboré, I.**, 2013. "DHS WORKING PAPERS." .

Pence, E. and **Paymar, M.**, 1993. *Education groups for men who batter: The Duluth model*. Springer Publishing Company.

Peterman, A., **Palermo, T. M.**, **Handa, S.**, **Seidenfeld, D.**, and **Team, Z. C. G. P. E.**, 2018. "List randomization for soliciting experience of intimate partner violence: Application to the evaluation of Zambia's unconditional child grant program." *Health economics* 27(3), 622–628.

Pouliquen, V., **Glennerster, R.**, and **Murray, J.**, 2021. "The Media or the Message? Experimental Evidence on Mass Media and Modern Contraception Uptake in Burkina Faso." .

Raj, A., **Saggurti, N.**, **Balaiah, D.**, and **Silverman, J. G.**, 2009. "Prevalence of child marriage and its effect on fertility and fertility-control outcomes of young women in India: a cross-sectional, observational study." *The Lancet* 373(9678), 1883–1889.

Robinson, J., 2012. "Limited insurance within the household: Evidence from a field experiment in Kenya." *American Economic journal: Applied Economics* 4(4), 140–164.

Rosenzweig, M. R. and **Wolpin, K. I.**, 2000. "Natural "Natural Experiments" in Economics." *Journal of Economic Literature* 38(4), 827–874.

Rossi, P., 2019. "Strategic choices in polygamous households: Theory and evidence from Senegal." *The Review of Economic Studies* 86(3), 1332–1370.

Rossi, P. and **Rouanet, L.**, 2015. "Gender preferences in Africa: A comparative analysis of fertility choices." *World Development* 72, 326–345.

Schuirmann, D. J., 1987. "A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability." *Journal of Pharmacokinetics and Biopharmaceutics* 15(6), 657–680.

Sen, A., 1990. *Gender and Cooperative Conflicts*. New York: Oxford University Press.

Sherin, K. M., **Sinacore, J. M.**, **Li, X.-Q.**, **Zitter, R. E.**, and **Shakil, A.**, 1998. "HITS: a short domestic violence screening tool for use in a family practice setting." *FAMILY MEDICINE-KANSAS CITY-* 30, 508–512.

Simes, R. J., 1986. "An improved Bonferroni procedure for multiple tests of significance." *Biometrika* 73(3), 751–754.

Simonsen, M., **Skipper, L.**, and **Skipper, N.**, 2010. "Price sensitivity of demand for prescription drugs: Exploiting a regression kink design." .

Smith, M. D., 1994. "Enhancing the quality of survey data on violence against women: A feminist approach." *Gender & Society* 8(1), 109–127.

Steinmetz, S. K., 1977. "The battered husband syndrome." *Victimology* 2(3-4), 499–509.

Stets, J. E., 2012. *Domestic violence and control*. Springer Science & Business Media.

Straus, M., 1979. "Measuring intra family conflict and violence: The conflict Tactics Scale." *J Marriage Family* 41, 75–88.

Straus, M. A., 1990. "Injury and frequency of assault and the "representative sample fallacy" in measuring wife beating and child abuse." *Physical violence in American families: Risk factors and adaptations to violence in* 8(145), 75–91.

—, 1999. "The controversy over domestic violence by women." *Violence in intimate relationships* , 17–44.

Straus, M. A., **Hamby, S. L.**, **Boney-McCoy, S.**, and **Sugarman, D. B.**, 1996. "The revised conflict tactics scales (CTS2) development and preliminary psychometric data." *Journal of Family Issues* 17(3), 283–316.

Stromquist, N. P., 1999. "The theoretical and practical bases for empowerment." *Women, education, and empowerment: Pathways towards autonomy* , 13–22.

Tauchen, H. V., **Witte, A. D.**, and **Long, S. K.**, 1991. "Domestic violence: A nonrandom affair." *International Economic Review* , 491–511.

Thomas, D., 1994. "Like Father, like Son; Like Mother, like Daughter: Parental Resources and Child Height." *Journal of Human Resources* 29(4).

Tourangeau, R. and **Yan, T.**, 2007. "Sensitive questions in surveys." *Psychological bulletin* 133(5), 859.

Treibich, C. and **Lépine, A.**, 2016. "Estimating misreporting in sensitive health behaviors: Evidence from condom use of female sex workers in Senegal." *London School of Hygiene and Tropical Medicine Working Paper* .

Tsai, C.-l., 2017. *Solutions to Turnout Over-Reporting: What Is Out There, What Works, and Can We Do Better?* Ph.D. thesis, University of Essex.

—, 2019. "Statistical analysis of the item-count technique using Stata." *The Stata journal* 19(2), 390–434.

Tsuchiya, T., **Hirai, Y.**, and **Ono, S.**, 2007. "A study of the properties of the item count technique." *Public Opinion Quarterly* 71(2), 253–272.

Udry, C., 1996. "Gender, agricultural production, and the theory of the household." *journal of Political Economy* 104(5), 1010–1046.

Uthman, O. A., **Lawoko, S.**, and **Moradi, T.**, 2010. "Sex disparities in attitudes towards intimate partner violence against women in sub-Saharan Africa: a socio-ecological analysis." *BMC Public Health* 10(1), 1.

Vicente-Serrano, S. M., **Beguería, S.**, and **López-Moreno, J. I.**, 2010. "A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index." *journal of climate* 23(7), 1696–1718.

Washington, E. L., 2008. "Female socialization: how daughters affect their legislator fathers." *American Economic Review* 98(1), 311–32.

Yount, K. M., 2005. "Resources, family organization, and domestic violence against married women in Minya, Egypt." *Journal of Marriage and Family* 67(3), 579–596.

Yount, K. M. and **Li, L.**, 2009. "Women's "justification" of domestic violence in Egypt." *Journal of Marriage and Family* 71(5), 1125–1140.

Yount, K. M., **VanderEnde, K. E.**, **Dodell, S.**, and **Cheong, Y. F.**, 2016. "Measurement of women's agency in Egypt: A national validation study." *Social indicators research* 128(3), 1171–1192.

Zigerell, L. J., 2011. "You wouldn't like me when I'm angry: list experiment misreporting." *Social Science Quarterly* 92(2), 552–562.