Essays on Applied Microeconomics -The Impacts of Internal Displacement-

A THESIS SUBMITTED TO THE UNIVERSITY OF DUBLIN, TRINITY COLLEGE IN APPLICATION FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
BY

Laura Muñoz Blanco

Supervised by

Nicola Mastrorocco
&
Marvin Suesse

2023

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

I consent to the examiner retaining a copy of the thesis beyond the examining period, should they so wish (EU GDPR May 2018).

Laura Muñoz Blanco

Summary

This dissertation consists of three essays at the intersection of applied microeconomics and development economics. It tackles questions on the consequences of internal displacement on intra-household and intra-community behaviours, with a special focus on gender, social cohesion, and health outcomes.

Chapter 1 provides evidence that exposure to events that trigger population outflows leads to early marriage by young women, putting them on a poor-life development path. Exploiting a novel dataset and the plausibly exogenous occurrence of earthquakes within Indonesian provinces, I show that an earthquake raises the annual hazard of women marrying before age 18 by 44%, compared to non-exposed young women. Earthquakes' overall effect on women's marriage age masks substantial heterogeneity. The effects are larger for earthquake-induced migrant versus left-behind women. Obtaining informal insurance from marriage induced migrants marry earlier as a financial coping strategy: a marriage payment, increased labour return when the husband joins the household, and social integration in receiving communities. These mechanisms do not have a role for left-behind women. I find evidence that a supply shock drives this result. Large population outflows and school building destruction that led to a drop in schooling explain the results for left-behind women.

Chapter 2 studies the long-term impacts of large inflows of forcibly displaced persons on displaced-hosting social participation outcomes. I exploit the construction of reservoirs during the Spanish dictatorship (1936-1975), which forced thousands of people into displacement. I profit from the margin of whether a pre-dictatorship project in 1933 planned the closest reservoir to a municipality, its size and distance to implement an instrumental variable approach. For this purpose, I rely on a newly-collected historical panel dataset on forced displacement and social participation. The results show a long-term and sizable decrease in voter turnout and the number of associations created in host communities. Additionally, the number of the forcibly displaced population relative to the natives mitigates the impacts. I propose two mechanisms: a decrease in general and institutional trust. A reservoir impacted natives and forcibly displaced populations differently, leading to inter-group clashes

with long-lasting effects on between-group and institutional trust.

Chapter 3 examines the impacts of the inflows of internally displaced people (IDP) on polio incidence in host communities. To tackle this question, I use the mass displacement of the population from the conflict-affected Federally Administered Tribal Areas (F.A.T.A.) to other districts in Pakistan from 2008 to 2022. In a difference-in-differences approach, I compare the new polio cases between host and non-host districts before and after 2007. I exploit the spatial distribution of districts relative to the border of the pre-colonial region of *Pashtunistan* to define the host and non-host districts. I find that districts that received the IDP population increased the number of additional polio cases per 100,000 inhabitants by 40% over the mean incidence compared to non-host districts. There are three underlying mechanisms: overpopulated communities with low immunization rates, precarious health conditions, and the congestion of health services in host communities.

To my beloved family,
for giving me the wings to fly
and the roots of perseverance.

To my truly loved partner in life, Paulo,
for keeping together our heads in the cloud
and our feet on the ground.

Acknowledgements

Each person's journey is different. In my case, the PhD journey pushed me to places where I had not been before. These places have been depression, desperation, anxiety, empowerment, consciousness, determination, and cheerfulness. Contradictorily, this journey has been an intellectual and psychological challenge but a stimulating and enjoyable adventure. My journey has been shaped by experiences, failures, opportunities, choices, and achievements. Nonetheless, this journey would indeed have never ended without the meaningful support, pure inspiration and unconditional love from many. I sincerely thank all of you.

I am deeply grateful to my supervisor, Prof. Nicola Mastrorocco, for your guidance and support throughout the PhD. Mainly, I thank you for helping me become a better researcher and project myself within Academia. I would like to also express my gratitude to Prof. Marvin Suesse for your constructive advice over the years and for supervising me during the last year with joy and patience. I am incredibly grateful to the two examiners for their comments on the thesis and the discussion during my Viva voce –Prof. Micaela Sviatschi and Prof. Selim Gulesci.

I thank Prof. Gaia Narciso for being a role model for me and my research. Thank you so much for all these years. On many occasions, you pushed up my determination and self-confidence. I will always be in debt with you. I express my sincerest gratitude to Prof. Selim Gulesci for your always insightful comments and time generosity. Our discussions always brought me peace of mind. I also want to thank Prof. Alejandra Ramos for supporting me since day one. You are a reference for me. I extend my sincerest gratitude to all members of the Economics department. In particular to Profs. Andrea Guariso, Davide Romelli, Carol Newman, Joseph Kopecky, Martina Zanella, and Nicola Fontana.

I benefit from visiting periods at Carlos III University and Andes University. I thank Profs. Luigi Minale and Andres Moya for hosting me. For helpful comments, I would like to show my gratitude to all the participants at the seminars, workshops and conferences where I presented my research and the scholars who generously opened their doors to discuss my research with them. Infinitely, thank you.

I thank the European Economic Association (EEA) for the Young Economist

Award. I thank Trinity College Dublin for the financial support. I also thank Trinity Research in Social Science (TRiSS) for the two fellowships I was awarded and for making TRiSS a fun and safe workspace.

My PhD fellows made a difference during this journey. I especially want to thank Anu Puthenmadathil, Federico Frattini and Juan David Duran for the countless support and unforgettable moments. Lastly, I would like to thank Dora Tuda, Luciano Ayala-Cantu, Bruno Morando, Gayane Vardanyan, Bryan Coyne, Jan-Luca Hennig, Louisa Roos, Iris Wohnsiedler, Eugenia Frezza, Naomi Crowther, Michael Mcrae, Maximilian Gunnewig, Simone Arrigoni, Matteo Prograxha, Giorgia Conte, Flavia Pellegrini, Danish Us-Salam, and Paul Blanchard.

The love and encouragement I have received from my partner and family are the genuine sources of this dissertation. I thank my family from deep in my heart for always holding me and my dreams. Thank you, "mamá", for teaching me to fight learning from adversity. To my sister and brother, for making my world special. You are my past, my present and my future. Thanks, "papá" and Paloma, for your support and cheering words. I want to thank my sister-in-law, Irene, and family-in-law for encouraging me during these years.

I want to thank my life partner and husband, Paulo, for making this journey with me. I deeply admire you for your optimistic and fighting spirit. Thank you so much for dreaming and laughing heartily with me. But especially for pushing my dreams, advocating my efforts, and supporting my decisions.

Lastly, a gigantic thank you to my friends for bringing poetry and reality to my life. In particular, thanks to Marcos for helping me love research, Alex and Eva for our endless conversations in Dublin, Alice for showing me an unknown path, and "las amapolas" for being part of my roots.

Contents

List of Tables

List of Figures

Introduction

More than 100 million people have been forced to leave their homes due to conflict, natural or human-made disasters. This means 1 in every 78 people on Earth is forcibly displaced (asylum seekers, refugees or internally displaced persons) and is projected to continue growing. Among them, 60% stay in their own countries, being *internally displaced persons* (or IDPs) (UNHCR 2022).

Internal displacement is a complex issue to address. It is often politically and logistically challenging to assist IDPs. Most do not live in camps but are dispersed among local communities, making identifying IDP populations and their needs challenging. The overwhelming majority of IDPs are women and children who are especially at risk of abuse of their fundamental rights. Context-specific factors can significantly affect the success of interventions for IDPs. These include the capacity and willingness of national and local institutions to assist IDPs (OCHA 2022). Overall, the special features of internal displacement compared to other forcibly displaced populations can differently shape its impacts on the internally displaced population, their host communities and the population left behind.

This thesis contains three chapters on the effects of internal displacement on intra-household and intra-community behaviours. The economic impacts of forced displacement are well established by the literature (Ibáñez, Rozo, and Urbina 2021; Alix-Garcia, Walker, et al. 2018; Fasani, Frattini, and Minale 2022; Arbatli and Gokmen 2023). There is limited evidence, however, about three main aspects: 1) the displacement impacts on gender roles and children's vulnerability within the household, 2) the long-term persistence of social-cohesion effects, and 3) what kind of interventions can mitigate the previously mentioned impacts.

Throughout this thesis, I address questions related to gender, social cohesion and children's health. To answer them, I employ cutting-edge quasi-natural experiments, exploit geo-referenced data to identify the push factor of the displacement and use microdata to characterise IDPs, their host or sending communities.

Chapter 1, "**Shifting Marriage Timing for Women: Destructive Events and Forced Displacement**", asks whether internal displacement changes the preferences for early marriage of women in sending communities and IDP women. I show that exposing girls to events that trigger large forcibly displaced population outflows leads to early marriage. The impacts hold for women in the sending communities and those forcibly displaced within their own country. However, IDP women experience the biggest impacts. To generate this evidence, I exploit the within-province and within-cohort plausibly exogenous occurrence of earthquakes in Indonesia from 1994 to 2014.

Cultural marriage norms and social network benefits explain the heterogeneous effects between women in sending communities and IDP women. I develop a marriage market model to rationalise the mechanisms behind the large heterogeneity. The model captures how forced displacement exacerbates the poor economic outcomes already impacted by earthquakes. While the marriage transfer from the groom to the bride (bride price) and the groom's labour contribution in matrilocal households affect earthquake-induced migrants' marriage outcomes to a certain extent, the network utility households can obtain from their offspring marriage is critical. I show supporting empirical evidence for the three mechanisms as coping strategies against their displacement. Finally, I show how unconditional cash transfer programs can mitigate the marriage effects for earthquake-induced migrant women.

These findings are important as they highlight that although women in sending communities and IDP women respond to different incentives, displacement adversely affects both. Understanding the economic incentives of each population is crucial for public policy to reduce early marriage and mitigate its welfare implication. These results also speak to the key channel of culture in shaping displaced populations' economic behaviour. Understanding the role of cultural norms can contribute to effective policy design and evaluation. If culture-specific factors affect parental incentives to use their daughters' marriage to cope with shocks, then culture-specific policies may be needed to target these incentives.

While the first chapter provides evidence of the marriage impacts of internal displacement on women at the origin and IDP women, in Chapters 2 and 3, I change the scope of the analysis by looking at the host communities' consequences. Chapter 2, "**Reservoir-induced displacement and social participation: Evidence from the Spanish Dictatorship**", examines the social cohesion effects between IDPs and native populations in host communities and how much the effects persist in the long-term. I provide new evidence of a long-term and sizable decrease in social participation in host communities. I measure social participation as voter turnout (in the general and municipal elections) and the number of non-profit associations. I show that the number of the forcibly displaced population relative to the natives mitigates the impacts. I generate this evidence by exploiting the construction of reservoirs in the Ebro's River catchment area during the Spanish dictatorship

(1936-1975), which forced thousands of people into displacement. I profit from the margin of whether the pre-dictatorship project of 1933 planned the closest reservoir to a municipality, the reservoir's size and distance to implement an instrumental variable approach. For this purpose, I rely on a newly-collected historical panel dataset on forced displacement and social participation from 1977 to 2018.

I argue that the decline in institutional and general trust is the underlying mechanism behind the main results. First, voters accounted for the effectiveness of the government responsible for the construction with two opposing behaviours: whereas the natives rewarded the government of Franco, the internally displaced population punished it. Their gratitude and displeasure have persisted decades later in their actions. Second, a reservoir impacted natives and forcibly displaced populations differently, leading to inter-group clashes with long-lasting effects on a decrease in between-group trust.

Overall, these results help to highlight the importance of responding to short-distance migration waves by increasing the inter-group cohesion between native and forcibly displaced populations. Neglecting to respond to forcibly displaced population's integration may end up hurting social participation in host communities, with long-lasting consequences over the following decades, and ultimately negatively impacting welfare.

Chapter 3, "**War on Polio Eradication: Reaching the Hard-to-Reach**", looks at how a forced population movement within a country could be a substantial barrier to disease eradication. I show suggesting evidence that the inflows of IDPs increase the polio incidence in host communities. To tackle this research question, I exploit the mass displacement of the population from the conflict-affected Federally Administered Tribal Areas (F.A.T.A.) to other districts in Pakistan. In a difference-in-differences approach, I compare the new polio cases between host and non-host districts before and after 2007. Due to cultural and linguistic barriers, most IDPs settled within the pre-colonial region of *Pashtunistan*. Therefore, to define the host and non-host districts, I use the spatial distribution of districts with respect to the border from *Pashtunistan*. I show that the districts that received the IDP population increased the number of new polio cases per 100,000 inhabitants.

I propose three potential mechanisms by which IDP inflows could slow polio eradication in host communities: a sudden increase in the population in communities with low vaccination rates, the precarious health conditions in host communities, and the congestion of health services in host communities. Although I can not disentangle the specific role of IDPs in polio transmission, I present descriptive evidence in line with the proposed mechanisms. Furthermore, I document that the arrival of non-vaccinated IDP children seems not to drive the increase in polio incidence. A vaccination program targeting forcibly displaced populations could be the policy behind these findings.

Three policy implications emerge from this chapter. First, reaching the hard-to-reach- such as IDP children or children in conflict zones- should be a public priority. Second, poor communities are the host communities of most of the IDPs. An effort to better integrate the IDP population into the health services and labour market should be made to improve the conditions in which they live. Finally, the inflow of new population comes with increased demand for health services. Even if the increase in the demand is modest, in locations where the health delivery or capacity is weak, it can congest the local health services. It is essential to reinforce host communities' health workforce and infrastructure, so locals and newcomers can access health services equally.

The major contribution of this thesis is to generate evidence on how internal displacement affects the socioeconomic dynamics within households and communities. This thesis primarily speaks to the literature on the consequences of forced displacement for displaced populations (Castells-Quintana, Lopez-Uribe, and McDermott 2022; Bahar, Ibanez, and Rozo 2021; Nakamura, Sigurdsson, and Steinsson 2022; Chyn 2018; Sacerdote 2012), their communities of origin (Bertoli, Gautrain, and Elie Murard forthcoming; Arbatli and Gokmen 2023; Testa 2021; Acemoglu, Hassan, and J. Robinson 2011; Engel and Ibáñez 2007)) and host locations (Arbatli and Gokmen 2023; E. Murard and Sakalli 2021; Bharadwaja and Mirzab 2019; Alix-Garcia, Walker, et al. 2018; Morales 2018). I add to this research agenda by presenting how a short-distance displacement is a determinant of early marriage for sending and IDP women (Chapter 1), a source of a long-standing decrease in social participation in host communities (Chapter 2), and a barrier to disease eradication (Chapter 3). On top of that, this thesis helps to shed light on the different push factors that generate forced displacement globally: natural disasters (Chapter 1), human-made disasters (Chapter 2), and conflicts (Chapter 3). Last but not least, there is a thematic contribution to three leading fields. First, Chapter 1 belongs to the literature on the marriage markets in developing countries. Much of the previous literature has focused on income shocks, legal enforcement and cultural norms (McGavock 2021; Corno, Hildebrandt, and Voena 2020; Chiappori, Salanié, and Weiss 2017; Greenwood, Guner, and Vandenbroucke 2017; Banerjee et al. 2013). However, no evidence exists of how events triggering population outflow affect women's marriage. Second, Chapter 2 contributes to the literature on long-term social cohesion determinants (Abel 2019; Levy 2018; Cagé and Rueda 2016). Finally, most of the existing literature on the determinants of disease eradication has focused on studying the mistrust of vaccines, the role of trade, public transportation closure, and the effect of refugees as a transmitting channel (Ibáñez, Moya, et al. 2023; Adda 2016; Oster 2012; Baez 2011). Chapter 3 adds to this research agenda by showing how the mass arrival of internally displaced populations can affect the incidence of polio in host communities if immunisation and health services are not reinforced.

Chapter 1

Shifting Marriage Timing for Women: Destructive Events and Forced Displacement

2022 European Economic Association Young Economist Award

1.1 Introduction

Marriage at an early age is associated with entrenched poverty and gender inequalities. It affects women disproportionately and has been linked to poor education, economic and health outcomes for both women and their children (Corno, Hildebrandt, and Voena 2020; Vogl 2013; Tertilt 2005). Despite its huge costs, early marriage remains widespread in developing countries. Women in conflict-affected and disaster-prone countries are particularly vulnerable. Indeed, shocks such as conflicts and disasters drop household financial capacity and have already relocated 100 million individuals worldwide (UNHCR 2022).

Growing attention has been devoted to studying early marriage (Corno, Hildebrandt, and Voena 2020; Bau 2021; Ashraf et al. 2020). Nonetheless, many of its determinants are still understudied. This paper contributes to closing this gap. I investigate whether shocks that trigger population outflows impact the age at marriage for young women.

What is the impact of shocks that trigger population outflows on the age at marriage for women? Responding to this question presents several challenges. We need data to follow women until they get married. And, omitted variables may encapsulates the endogeneity problem. Hence, to study this question, I exploit a novel data set to track women until they marry alongside an exogenous shock to the marriage market. In particular, I profit from the plausibly exogenous variation in the timing and geographical occurrence of earthquakes within Indonesian provinces from 1994 to 2014. The threat goes beyond women from communities in disaster-affected areas. It also affects women that migrate induced by a disaster. Hence, I can evaluate if marriage responses differ for migrants and those left behind in earthquake-affected areas.

I identify earthquakes' effects on marriage timing by comparing young women from the same cohort, age and province who are exposed to earthquakes and those not exposed. I combine two useful sources of variation: i) within-province time and geographical variation in the plausibly exogenous occurrence of earthquakes and ii) within-cohort variation in the age of exposure to an earthquake. Using within-province and within-year-of-birth variation in the occurrence of earthquakes and marriage, I implement a difference-in-differences strategy in a hazard model of the marriage market. I then identify the effect of earthquakes on marriage patterns for *earthquake-induced migrant women* and *left-behind women*. To compare them, I rely on a third source of variation: within-earthquake household variation in the exposure to that earthquake that leads to changes in migration decisions.

I build a geolocalised person-age panel dataset to implement my identification strategy. The dataset combines information on marriage, migration and household characteristics (education, ethnicity, parents' and spouse's attributes, among others) with satellite image information on the occurrence of earthquakes (the date, their epicentre, and area of exposure). Using these data, I can identify households affected by an earthquake at any point in time, and the intensity of exposure to the earthquake and follow them over time.

I generate three main findings. First, earthquakes increase the incidence of female marriage below 18. Second, the overall effect masks important heterogeneity: the effect on marriage is stronger when women migrate after the disaster than those who stay behind. Third, I document that bringing ahead women's marriage can be a vital financial mechanism against a migration shock, which explains the heterogeneous effect for *earthquake-induced migrant women*.

I show that an earthquake raises the annual marriage hazard by 19% compared to young women from the same cohort, age, and province but not exposed to earthquakes. The effects are sizable and chiefly affect marriage below 18: an earthquake raises the annual hazard of marriage between ages 12 and 17 by 44%. I provide evidence supporting the parallel trends assumption using an event study specification that allows the relative effect of earthquakes on exposed and non-exposed young women to vary over time.[1]

After establishing that earthquakes have large effects on the timing of marriage, I examine an important implication of destructive events: disaster-induced migration.

[1] These findings are robust to a broad set of alternative definitions of a destructive earthquake and a range of difference-in-differences estimators.

Earthquakes cause a 49% increase in migration compared to the baseline mean. The probability of migrating depends positively on the vibration of the ground during an earthquake and negatively on the distance to the epicentre.[2]

To study whether differential patterns in marriage exist, I examine heterogeneity between *earthquake-induced migrants* and those *those left-behind* in earthquake-affected areas. I show that migration induced by an earthquake is associated with a 72% increase in the annual probability of getting married compared to *left-behind women*. I also incorporate family fixed effects into the design to sharpen the identification and use a sister pair comparison to show the effect. The findings hold. The heterogeneous results for *earthquake-induced migrant* respond to the differential effects of earthquakes between *earthquake-induced migrant women* and *left-behind women*. While both are affected by earthquakes, the *earthquake-induced migration* exacerbates the poor economic outcomes already affected by an earthquake. On top of that, migrant households end up in a new marriage market in which they have no local networks to turn to.

There are three forms of informal insurance that *earthquake-induced migrant women* can obtain from marrying. First, the bride's parents can benefit from a marriage payment from the groom's family (bride price). I take advantage of within-country variation in the traditional practice of bride price across ethnic groups to evaluate whether receiving a marriage transfer changes the results.[3] The effects of earthquakes on the annual marriage hazard is 71% larger among young women from ethnic groups that traditionally practice bride price with respect to women that do not traditionally practice the custom. A transfer at marriage can alleviate women's household financial constraints after migrating. Second, linked to the matrilocality tradition, the aggregate labour return of the woman's household increases when newly formed couples join the household at the moment of marriage. Earthquake impact is seven-time stronger for women in ethnic groups that traditionally practices matrilocality compared to those that do not traditionally practice the custom. This additional economic return can help smooth the shock of migrating. Third, marriage can also serve as a means of facilitating social integration in receiving communities and boost their socioeconomic network. I use data on involvement in community organizations to assess the integration channel. Participation in communal groups decreases marriage effects, corresponding to 12% of the baseline results. This finding suggests that women whose households are already well integrated at the new destination are less likely to marry earlier. Thus, the opposite would happen when households are poorly integrated.

[2] Table 1.2 shows that migration motivated by marriage does not threaten the identification in this context.

[3] Relative to traditional practices, modern data are more likely to be endogenous to modern factors, including earthquakes or migration induced by the disaster.

I develop a simple marriage market model to help reconcile the empirical findings on the mechanisms for *earthquake-induced migrant women.* My theoretical framework models the relationship between marriage market and a disaster-induced migration. In particular, I extend a model, originally developed by (Corno, Hildebrandt, and Voena 2020), by adding three new features: 1) households are matrilocal, 2) marriage markets are in receiving communities, 3) households acquire new networks with their offspring marriage. While bride price and groom´s labor contribution affects marriage outcomes at a certain extend, the network utility that households can obtain from their offspring marriage is the critical component.

Instead, I find no evidence of informal insurance mechanisms being present for *left-behind women.* I find evidence that the effects for *left-behind women* have to do with two main mechanisms. First, earthquakes trigger large population outflows that, even if they do not affect the sex ratio, increase women's preference for an early marriage rather than waiting and risking not finding a good match in the future. Second, I provide evidence that school buildings destruction due to earthquakes leads to a drop in schooling and, consequently, an increase in marriages below the age of 18. These findings support the idea that *left-behind women* do not marry earlier to benefit from consumption smoothing mechanisms because markets in earthquake-affected areas are credit-constrained in the aftermath of a disaster.

Although marrying earlier is a financial coping strategy for a *disaster-induced migrant's families*, a marriage at an earlier age has tremendous welfare implications for women and their households. By comparing married and unmarried migrant women from the same cohort, age and province, I show that married women have their first child earlier and have less likelihood of being employed. Migration also affects the characteristics of marriages (increases education gap, decreases polygynous marriage). I also find that the consumption capacity of *disaster-induced migrant's households* do not change with the early marriage of their daughter. These findings imply that such households are not better off after their daughter's marriage.

Having provided evidence that *earthquake-induced migrant* women marry earlier to cope with the negative consequences of migration, I then analyse how policy can change household incentives by addressing the underlying mechanisms. In the last part of the paper, I provide evidence of one potential measure: Unconditional Cash Transfers (UCT). I show how UCT mitigates the marriage effects for *earthquake-induced migrant women.*

My findings have several potential policy implications. First, given the growing number of crisis-driven out-migration flows, governments are increasingly focusing on ways to combat its consequences. There has been a growing consensus on the adverse effects of population outflows on education and labour outcomes (S. Becker and Ferrara 2019). I contribute to the current political debate by generating evidence on how exposure to shocks triggering population outflows early in life can explain female early marriage. Second, by documenting the differential impacts for disaster-induced migrants, this paper helps uncover which policies could effectively reduce the incidence of early marriage. Since effects are driven by exposure during childhood or young adulthood, policies targeting early marriage incentives can reduce its welfare implications and can potentially offer a more cost-effective way to respond to their migration. In addition, my results suggest that considering the ages when disaster-induced migrant women are most vulnerable to marriage is essential for future policies' better design. For example, a future cash transfer program aimed at decreasing early marriage after a displacement would be more efficient if targeted, not only below 18, but until 22 years old. Third, my findings highlight the importance of culture in shaping displaced populations' economic behaviour. Understanding the role of cultural norms can contribute to effective policy design and evaluation. If culture-specific factors affect parental incentives to use their daughters' marriage to cope with shocks, then culture-specific policies may be needed to target these incentives.

This paper speaks to three strands of the literature. First, this paper belongs to the literature on the determinants of marriage markets in developing countries (Banerjee et al. 2013; Chiappori, Salanié, and Weiss 2017; Corno, Hildebrandt, and Voena 2020; McGavock 2021; Greenwood, Guner, and Vandenbroucke 2017). (1) Much of the previous literature has focused on income shocks, legal enforcement and cultural norms. However, there is no evidence of how shocks triggering population outflow affect female marriage for *leave-behind* versus migrant women. Carlana and M. Tabellini 2020 is the closest related paper. The study examines the effects of immigration on *natives*' marriage across US cities between 1910 and 1930. (2) Moreover, (Corno, Hildebrandt, and Voena 2020) have analysed the effect of droughts - as a proxy for income shocks- on the timing of marriage. In contrast to theirs, my paper focuses on a distinct setting and its effects. First, whereas droughts are weather-related events, earthquakes are seismic events, making them unpredictable. Second, while droughts also affect migration decisions, Corno, Hildebrandt, and Voena 2020 do not look at the migration dimension. Third, one of the limitations to run migration analysis is data requirements. My dataset allows me to track individuals over time and across space. (3) A related literature examines whether traditional norms determine women's marriage (Corno, Hildebrandt, and Voena 2020), education (Ashraf et al. 2020; Bau 2021); or health (Bhalotraa, Chakravartyb, and Gulesci 2020) outcomes. I complement this literature by showing that cultural norms at marriage are a key part of understanding the differential effects on age at marriage between migrants and *left-behind young women. Disaster-induced migrants* gain informal financial coping strategies from their marriage.

Second, I contribute to an emerging literature on the consequences of forced displacement for the population left behind (a synthesis of this literature is provided in S. Becker and Ferrara 2019) and migrants themselves (Nakamura, Sigurdsson, and Steinsson 2022; Chyn 2018; Sacerdote 2012; Bahar, Ibanez, and Rozo 2021; Castells-Quintana, Lopez-Uribe, and McDermott 2022). Most empirical studies in this area have analysed education (Chiovelli et al. 2021), economic (Fasani, Frattini, and Minale 2022), or political (Rozo and Vargas 2021) outcomes. This paper focuses on marriage outcomes (Lu, Siddiqui, and Bharadwaj 2021). I find that the effects for *disaster-induced migrants* are larger than for *left-behind women*. While earthquakes affect both, *migrants* end up in a new marriage market in which they have no local networks to turn to. *Migrants* also suffer an additional income shock which subsequently affects the timing of marriage. In line with (Nakamura, Sigurdsson, and Steinsson 2022; Chiovelli et al. 2021), I provide empirical evidence of how migration is translated into an income shock. While most of this literature has explored cross-country migration, I complement this literature by studying short-distance migration (internal displacement). Furthermore, my rich dataset allows me to dig into the potential mechanisms behind the heterogeneous results and evaluate a cash transfers program as a potential policy to mitigate the effects for *disaster-induced migrants* (Özler et al. 2021).

Third, my results complement the literature studying households' responses to natural disasters (Gunnsteinsson et al. 2022; Deryugina, Kawano, and Levitt 2018; Hanaoka, Shigeoka, and Watanabe 2018; Kirchberger 2017; Gignoux and Menéndez 2016). Part of this literature emphasises the supply side, and explores the impacts of natural disasters on labour markets (e.g, Deryugina, Kawano, and Levitt 2018; Kirchberger 2017; Gignoux and Menéndez 2016), health and risk-preferences outcomes (Gunnsteinsson et al. 2022; Hanaoka, Shigeoka, and Watanabe 2018). "An average of 25.3 million displacements have been brought on each year since 2008 by natural disasters alone" (IDMC 2020). Few empirical papers investigate how the effects of natural disasters can potentially differ when households migrate or are left behind in disaster-affected areas. This paper adds to the literature by investigating the heterogeneous effects of destructive events on age at marriage for women between *disaster-induced migrant women* and those *left-behind*.

The remainder of the paper is organised as follows. Section 2 introduces the setting. Section 3 summarises the data. Section 4 describes the empirical strategy. Sections 5 and 6 show the results and mechanisms. Section 7 examines the effects of a specific policy targeting the underlying mechanisms. I present the robustness checks in Section 8, followed by a conclusion section.

1.2 Context

In this section, I provide background information relevant to my analysis. First, I describe Indonesia's marriage market and the traditional marriage norms. Second, I present an overview of earthquakes in Indonesia and describe the set of disasters I exploit for identification. This paper focuses on changes within provinces in the exposure to earthquakes from 1994 to 2014.

1.2.1 Marriage Market in Indonesia

In Indonesia 81% of women married by 23 versus 51% of men. And, fewer than four percent of women over the age of 40 have never married (Jones 2004). Ethnicity and religion are crucial in the marriage process. 1 in 10 and only a 0.5% of married couples have different ethnicity or faiths, respectively. (Indonesia Population Census 2010).[4]

The female timing of marriage has tremendous welfare implications for women and their children. Marriage outcomes are an important component of the returns to education, especially for women (e.g., Chiappori, Salanié, and Weiss 2017; Ashraf et al. 2020). Marriage decisions affect early fertility and economic and health outcomes for women and their children (Corno, Hildebrandt, and Voena 2020; Vogl 2013; Tertilt 2005).

In 2020, one in nine married women in Indonesia were married before 18. In the past, marriage below 18 was a common practice on the archipelago (UNICEF 2020). Legal enforcement can not explain entirely the decrease in child marriage (below 18) practice. Until 2019 the legal age at marriage for women was 16. Nonetheless, islamic law and the adat laws (customary laws) in many parts of Indonesia allow the marriage below 18 (Nisa 2016). The practice of child marriage is not conditional on socioeconomic level or religion. Although, there is variation across regions and a higher prevalence in rural areas (SUSENAS 2015). Child marriage is also practised by the higher classes of the country regardless the faith.

1.2.2 Marriage Norms Variation in Indonesia

Local cultural norms are crucial for economic development (Ashraf et al. 2020). Indonesia is a country that has sub-national variation in the practice of marriage norms across ethnic groups. In Indonesia exists more than 300 different ethnic groups, each of which follows different traditional practices at the moment of marriage (i.e.

[4] It needs to be interpreted carefully because of the common practice of premarital conversion. Indonesia is not an Islamic state, but 86.7% of its population is Muslim (Indonesia Population Census 2010).

bride payments, kinship, polygyny and matrilineality) (K. Robinson 2010). Cultural norms in Indonesia still persist today and influence the socio-economic decisions of an entire household.

The transfer of a payment at the time of marriage is a custom that has deep historical roots (Ashraf et al. 2020).[5] In particular, a marriage payment from the groom´s family to the bride (bride price) is widely practised in Indonesia, with a within-country variation in the practice. Importantly, none of the ethnic groups within Indonesia traditionally practice dowry (a transfer from the bride and/or her family to the groom's parents upon marriage). And, it is common for the value of the bride price to be more than a year's income (Anderson 2007).

In Indonesia, there is as well an ethnicity-level variation in the practice of traditional kinship in Indonesia.[6] Kinship tradition determines daughters' and sons' post-marriage residences. In this paper, I focus on variation when husband joins wife´s household (matrilocality tradition) or not (patrilocality or neolocality tradition).[7]. Kinship traditions are decisive for economic outcomes, particularly in low income countries (Bau 2021).

In this paper, I focus on the role of marriage payments and post-marriage residency traditions. Nonetheless, I also evaluate the potential role of other cultural customs (i.e., polygyny (men have more than one wife), matrilineality (female participation in agriculture), arranged marriage and marriage migration) on the timing of marriage. But, arranged marriage and marriage migration are non-common phenomenon in Indonesia.

1.2.3 Earthquakes in Indonesia

Nearly 45 per cent of the world's natural disasters occur in the Asia-Pacific region (UNFPA 2018). Being located on the Pacific Ring of Fire makes Indonesia one of the world's most natural disaster-prone countries.

[5] There are three dominant theories in anthropology to explain the origin of the bride price practice are: The first is that the custom historically originated in patrilineal societies, where the wife joins the husband's kinship group following marriage (Vroklage 1952). The second theory links the practice of bride price to the participation of women in agriculture (Boserup 1970). The last, related factor that is potentially relevant for bride price is the practice of polygyny (Grossbard 1978; G. Becker 1973; Tertilt 2005)

[6] There are different theories on the origin of matrilocality traditions. One theory argues that early hunter-gatherer societies were typically matrilineal (lineage and inheritance pass through the mother's line, and a son usually inherits from his maternal uncle) because sexual promiscuity made it challenging to identify a child's father (Engels, 1942). An alternative theory is that matrilineality tends to occur in horticultural societies where women often have a more dominant role in agriculture (Jones, 2011). Finally, some anthropologists have also linked matrilineality to dowry and patrilineality to bride price (Vroklage, 1952).

[7] Whereas in patrilocality tradition wife goes to live to his husband´s household, in neolocality newly form couples creates a new household after marriage

Indonesia is the country with the highest exposure to big earthquakes (USGS). Earthquakes are probably the biggest threat in terms of natural disasters in Indonesia as they come suddenly and cannot be predicted. Earthquakes occur across the entire country, affecting also populous areas. On average, Indonesia experiences about one earthquake with a magnitude of six or higher in the Richer scale per year (Indonesia-Investments). It implies that 62.4% of the total Indonesian population is exposed to destructive earthquakes yearly (UNFPA 2018).

Earthquakes drop the local economy (Kirchberger 2017; Gignoux and Menéndez 2016) by destroying houses and public amenities. But, earthquakes also push millions of persons to leave their places of resident. Precisely, earthquakes led to 60% of the new disaster-displacements worldwide in 2019 (IDMC 2020).In the areas affected by earthquakes, the large population outflow trigger by the shock changes the demographic composition of marriage markets. In particular, quick migration decisions led to poor economic outcomes. And, forcibly displaced population end up in a new marriage market at destination, with a lack of local networks that could help them to cope with their migration shock.

This paper investigates the impacts of shocks that trigger population outflows on marriage decisions for young women. I focus on large earthquakes to ensure that I capture the out-migration dimension. In line with (Gignoux and Menéndez 2016), I define large earthquake as an earthquake with a ground shaking of at least VII in some of its locations affected.[8] Therefore, this paper exploits the within-province time variation in the plausibly exogenous occurrence of destructive earthquakes in Indonesia from 1994 to 2014. Figure 1.1 shows the variation in destructive earthquakes across time and space in Indonesia that this paper uses.

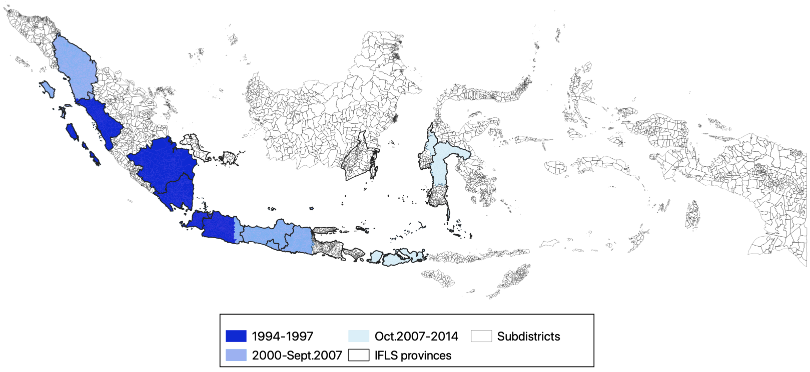
1.3 Data

This paper uses three main datasets that provide individual-level variation across geographic region and overtime. The first two datasets-individual longitudinal and earthquake exposure data- provide the tools to construct the outcome of interest and treatment variable. Finally, the ethnicity-level data provide information on variation in the traditional practice of cultural norms across ethnic groups. By using historical information, I circumvent potential endogenous problems with current engagement in marriage norms.

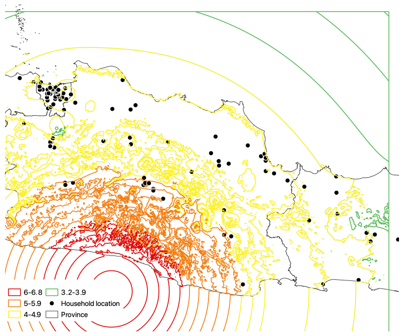
[8] While an earthquake has only one magnitude and one epicentre; it produces a range of ground shaking in the surrounding area. For example, the Richter scale measures its magnitude at its epicenter. The modified Mercalli scale or the Rossi-Forrel scale are commonly used to measure the ground shaking of an earthquake in the vicinity of the epicenter.

Figure 1.1: Large earthquakes in Indonesia (1994-2014)

a) Variation across years and provinces



b) Variation in earthquake ground shaking across households locations



Note: This figure shows this paper's sources of variation from USG Survey: large earthquakes (with an intensity of at least VII in some of its locations affected (Gignoux and Menéndez 2016) in Indonesia from 1994 to 2014. Figure a presents the variation in the occurrence of earthquakes across survey years and provinces in my sample. However, I profit from monthly and sub-district variation within a province for my identification. Darker colours correspond to earlier occurrences. Figure b shows the household variation in the intensity of exposure to the same earthquake (e.g. ground shaking or Modified Mercalli).

1.3.1 Longitudinal Individual Data

I track women until they actually get marry by using the Indonesia Family Life Survey (IFLS) for 1997, 2000, 2007 and 2014.

The IFLS is an ongoing longitudinal survey at the individual level.[9] Subsequent waves re-interviewed the original households (and all the members older than 15) and tracked individuals who had moved to another destination within the country (Strauss, Witoelar, and Sikoki 2016).[10]

[9] Five waves have been conducted so far: 1993 (IFLS1), 1997–1998 (IFLS2), 2000 (IFLS3) and 2007–2008 (IFLS4), 2014-2015 (IFLS5). The IFLS sampling scheme is stratified on 13 of the 27 provinces of the country and urban/rural locations and then randomly sampled within these strata. Together these provinces encompass approximately 83 per cent of the Indonesian population and much of its heterogeneity.

[10] The IFLS is well-known by its low attrition rates, 86.9% are interviewed in all five rounds (higher than most longitudinal surveys in the US and Europe). Attrition rates do not differ between population affected and non-affected by earthquakes (Strauss, Witoelar, and Sikoki 2016). The IFLS also gathers information on international migration

The IFLS provides retrospective data on migration, marriages and covariates at the individual level. For instance, the IFLS gathers information on education, labour, ethnicity, religiosity, family and spouses characteristics, among other (Frankenberg and Thomas 2000). Timing and spatial data for each migration allow me to track individuals before and after an earthquake. This information is vital to define the sub-district of residence, and, sub-district of destination, in case of a migration. Furthermore, the GPS coordinates of the IFLS survey clusters enable to identify the coordinates of each household.

1.3.2 Earthquake Exposure Data

The temporal and geographic variation in the occurrence of destructive earthquakes that defines the changes in exposure to earthquakes is drawn from the United States Geological Survey (USGS). The observed information is at the daily and desa level. However, in order to match the IFLS information I focus on the variation at monthly and subdistrict level.[11] See Table A.1.1 with descriptive statistics of earthquakes.

Geocode satellite data from ShakeMaps-USGS allows me to perform a more disaggregated analysis at the grill-cell level. These data provide information on the ground shaking characteristics produced by an earthquake at each particular location. The ShakeMap shapefile uses the modified Mercalli intensity scale, an exogenous measure calculated based on peak ground velocity and peak ground acceleration (Wald et al. 2006). This scale consists of increasing levels of ground shaking that range from imperceptible shaking (zero) to catastrophic destruction (ten). Using this scale lets me capture the change in an earthquake's impact across households. Figure 1.1 shows a ground shaking satellite image for a destructive earthquake in Indonesia. There is a progressive variation in the earthquake ground shaking across each grill cell. This suggests that the modified Mercalli intensity scale is a reliable exogenous measure of earthquake exposure at the lowest administrative level of Indonesia: communities or *desas*.

Importantly for my identification, I take advantage of the fact that the IFLS collects the GPS coordinates of survey clusters by spatially linking data on the ground shaking of earthquakes with an individual's location (Kirchberger 2017; Gignoux and Menéndez 2016). Therefore, this approach enables me to capture the variation in the exposure to earthquakes across individuals.

[11] Indonesia's administrative levels are provinces (provinsi), which are divided into districts that are further divided into sub-districts (kecamatan). Districts can be cities (Kota) or regencies (Kabupaten). Sub-districts are divided into villages (desa). For the purposes of this study, cluster and community refer to the 321 original Indonesia Family Life Survey (IFLS) communities plus the communities of migrants households.

Conditional on certain locations, the impossibility of predicting when earthquakes occur makes them an exogenous event, which is a crucial element of my identification strategy. I profit from 21 earthquakes in Indonesia from 1994 to 2014 for the identification.

1.3.3 Traditional Norms Data

For my main analysis, I use historical measures of marriage norms, instead of contemporary measures from the IFLS. Relative to traditional practices, modern data are more likely to be endogenous to modern factors, including earthquakes or migration induced by the disaster. To examine if women engaged in marriage customs are more likely to change the timing of their first marriage, I use ethnicity-level data on traditional engagement in cultural norms, from Murdock 1967 Ethnographic Atlas. I start using transfers made at marriage and traditional kinship practice, categorising ethnic groups as engaged or not in each practice.

The Indonesian Family Life Survey (IFLS) collects the self-reported ethnicities of respondents, which I use to assign the presence of a traditional bride price custom (or not) to a married couple (Ashraf et al. 2020).[12] The prevalence of each marriage custom for Indonesia is reported in Table A.1.2. None of the ethnic groups within Indonesia traditionally practice dowry.

Information on actual bride price payments at marriage comes from the IFLS. [13] The IFLS also gathers information on the household of residence after marriage. I profit from these data to look at the contemporary bride price and kinship practice. Furthermore, I benefit from the fact that the Ethnographic Atlas captures information on the traditional presence of another custom (i.e., matrilineality and polygyny) to evaluate its role on marriage decisions.

1.3.4 Other Data

I profit from data from four different sources to measure over time changes in marriage market at aggregate level. I use Population census (years 1990, 2000, and 2010) to measure population density, sex ratio, and unmarried population at district level. I exploit the geo-referenced dataset on ethnic groups homeland from (Weidmann,

[12] I use the ethnicity of the survey respondent and respondent's parent to determine whether the woman belongs to a bride price ethnic group. In cases where both the husband and the wife were asked about the marriage, we use the husband's responses and ethnicity under the assumption that he is more likely to remember the bride price correctly. Since intermarriage between ethnic groups with different bride price customs is very low in Indonesia (1.5% in the Intercensal Survey data), this decision is not consequential

[13] The IFLS asks about dowry and bride price together and does not distinguish between the two. However, according to the IFLS documentation, the marriage custom is bride price except for marriages among the matrilocal Minangkabau group (Frankenberg and Karoly 1995), which I omit from the analysis.

Rød, and Cederman 2010) to measure the proximity to an ethnic homeland. To do so, I measure the distance from their ethnic homeland to their sub-district of residence. I take advantage from DMSP dataset to build a proxy of local economic development using night light intensity. I use data from DesInventar survey to capture earthquake destruction at district level.

1.3.5 Sample

The IFLS data contain 83,770 individuals in the last round. About 50.51% are women. From this sample, I keep 8,608 young women at least 23 years old at the last interview to avoid excluding never-married women (i.e.ages ranging from 23 to 34). By this age, 81% of women are marriage. Additionally, to ensure that women were interviewed during the IFLS time horizon, I focus on women born after 1980.[14] Appendix Table A.1.2 presents some descriptive statistics.

1.4 Empirical Strategy

1.4.1 Earthquakes Analysis on the Timing of Marriage

To estimate the effect of earthquakes on the timing of marriage, I exploit the plausibly exogenous timing variation in the occurrence of destructive earthquakes within Indonesian provinces in a hazard model, adapted from (Corno, Hildebrandt, and Voena 2020). In particular, I exploit the within-province exogenous variation in the timing of earthquakes to implement a difference-in-differences strategy in a hazard model. Figure 1.1 shows this variation.

In order to capture marriage behaviours age by age, I convert my dataset into person-age-month panel format. The duration of interest is the time between t_0, the age when a woman is first at risk of getting married, and the age when she marries for the first time t_m. In my analysis, t_0 is age 12, which is the minimum age at which a negligible number of women in my sample report getting married for the first time. See Appendix Figure A.1.1 for a visualization of the distribution of women's ages at first marriage. Hence, a woman who is married at age t_m contributes (t_m - t_0 + 1) observations to the sample. With one observation for each at-risk year until she is married, after which she exits the data.[15] To, later on, merge these individual data with earthquake data at the month level and covariates at the year level. Appendix Figure A.1.2 shows the dataset's structure with an example.

[14] Therefore, I can track women and their sociodemographic information from the age of 12 to the age of 23.

[15] For example, a woman who is married at age 20 would appear nine times in the regression for timing of marriage, and her marriage vector would be $M_{i,k,12}$, ..., $M_{i,k,19}$, $M_{i,k,20}$ = 0, ..., 0, 1. A woman who is not married by age 17 appears in the data six times, and her marriage vector is a string of zeroes.

Using this panel data and sample, I estimate the probability of marriage of woman i living in subdistrict s born in cohort k and entering her first marriage at age a as follows:

$$Y_{i,s,k,a} = \beta_0 + \beta_1 Eq_{s,p,a,m} + \beta_2 X_i + \alpha_p + \gamma_a + \delta_k + \zeta_u + \epsilon_{i,d} \tag{1.1}$$

The dependent variable, $Y_{i,s,k,a}$ is a binary variable coded as 1 in the year the woman gets married, and zero otherwise. Since I am interested in the timing of marriage rather than if women marry or not, I examine data on women until age 22. Thus, women married after age 22 are right censored. The variable $Eq_{s,p,a,m}$ is a time-varying measure of earthquake in sub-district **s** within province p occurring during the year in which the woman is age a and month m. Specifically, included in $Eq_{s,p,a,m}$ is a dummy indicator that switches to 1 for a earthquake occurring in sub-district of resident s in a given year a and month m, 0 otherwise. β_1 is the main coefficient of interest and measures the effect of earthquakes on the probability of marriage. I control for province-specific fixed effects, α_p, to account for time-invariant local unobservable characteristics, such as geographic, economic and cultural factors. γ_a is a vector of age fixed effects, which controls for the fact that marriage has a different probability to occur at different ages. Year-of-birth fixed effects ,δ_k, control for cohort effects, and, urban fixed effects ,ζ_u, account for different probability of marriage in urban places. I further control for a measure of individual level covariates measured a year before an earthquake, X_i (mother education and religion).[16] Standard errors are clustered at district level to allow for serial correlation in the error terms across women in the same area, and show robustness to clustering at lower and larger geographic units.[17].

With the inclusion of province and year of birth fixed effects, the impact of earthquakes on the marriage hazard is identified from within-province and within-year-of-birth variation in the occurrence of earthquake and marriage outcomes. The key identifying assumption of the analysis is that, within a given location and year of birth, earthquakes, $Eq_{s,p,a,m}$ are orthogonal to potential confounders. The exogeneity of earthquakes is crucial in my setting because, given my interest in looking also at the heterogeneity between *earthquake-induced migrants* and *non-migrants*, I eliminate potential anticipation to the shock.

I provide evidence supporting the parallel trends assumption by estimating an event study version of the baseline specification that allows the effects to vary over time. In particular, I estimate the following specification:

[16] I do not control for father education and number of siblings due to potential correlation with mother education. Appendix Table A.1.16 shows that results remain constant including both covariates.

[17] The IFLS sample is representative at province level. The variation decreases when clustering error terms at subdistrict level. But, results hold when clustering standard errors at sub-district level (Table A.1.15)

$$Y_{i,s,k,a} = \beta_0 + \sum_{p=-5}^{3} \beta_p Eq_{s,p,a,m} + \beta_2 X_i + \alpha_p + \gamma_a + \delta_k + \zeta_u + \epsilon_{i,d} \tag{1.2}$$

where variables are defined as above. To reduce noise, I constrain the effect to be constant by year since treatment.

Natural disasters trigger the vast majority of the forcibly displaced population within a country (Nakamura, Sigurdsson, and Steinsson 2022; IDMC 2020).[18] Earthquake's estimates from equation 1 capture all women exposed to earthquakes. Therefore, β_1 from equation 1 includes two types of women. First, women that stayed in the area affected by an earthquake. Second, a share of my sample migrates induced by the disaster as a strategy to adapt to the shock. As a further step I investigate if and how earthquakes affect left-behind young women differently. To do so, I exploit the plausibly exogenous geographic variation in the occurrence of earthquakes within Indonesian provinces. Particularly, I rely on the plausible exogenous variation in earthquake's ground shaking across households. The rest of my empirical strategy aims at investigating if earthquakes change migration decisions (section 4.2) and learn if the impacts of earthquakes differ between *earthquake-induced migrant* and *non-migrant* women (section 4.3).

1.4.2 Do Earthquakes Impact Migration?

In order to examine the migration consequences of earthquakes, I start estimating the impact of earthquakes on migration decisions for all individuals in the sample at the survey level. Equation 3 presents the specification:

$$Y_{i,s,y_s} = \beta_0 + \beta_1 Eq_{s,y_s} + \beta_2 Eq_{s,y_s} * X_i + \alpha_p + \gamma_{y_s} + \zeta_u + \epsilon_{i,d} \tag{1.3}$$

where Y_{i,s,y_s} is a binary variable coded as 1 if migrating outside the sub-district of resident after an earthquake, zero otherwise. The exposure to an earthquake, Eq_{s,y_s}, is a dummy indicator that switches to 1 for a earthquake occurring in sub-district of resident s in a given survey-year, 0 otherwise. I control for year survey-province fixed effects $\alpha_{y_s,i}$, age fixed effects, and urban fixed-effects, ζ_u. We may be concerned that earthquakes affect differently conditional on gender, religion, ethnicity or potential marriage. I include an interaction of the exposure to an earthquake, Eq_{s,y_s}, to individual level covariates before an earthquake strikes, X_i (women, non-javanese, non-muslim, age gap to 23 among women). Standard errors are clustered at district level. I also profit from the available information on marriage migration from the IFLS to assess if earthquakes affect migration decisions related to marriage.

[18] Forced displacement occurs when individuals have been obliged to leave their habitual residence as a result of or to avoid the effects of events such as armed conflict, generalized violence, human rights abuses, natural or man-made disasters, and/or development projects (UNHCR).

To ensure that the migration decisions are directly driven by an earthquake, I refine my migration definition. To do so, I exploit year and month information on (i) the IFLS interview, t_{IFLS}; (ii) earthquakes, t; (iii) and women's migration decisions, t_m; as well as, spacial data on (iv) sub-districts affected by an earthquake, s_e; (v) sub-district of residence of women i at t, t_m, and t_{m+1}. Therefore, woman i is classified as *earthquake-induced migrant* if she was in a sub-district affected by an earthquake, s_e, when it occurred, at time t, and, the timing of her migration, t_m, was immediately after time t, and her place of residence at t_{m+1} was within Indonesia.[19] In what follows, I call this migration definition as *earthquake-induced migration* (or forced displacement), and women that stay in earthquake-affected areas as *non-migrants*. I estimate the above-presented estimation with my *earthquake-induced migration* outcome.

When a destructive natural hazard occurs, the population affected is often relocated to shelters for a period that could range from 6 to 24 months. The concern is that those in the shelter may be self-reported as non-migrants even if they are. To overcome this limitation, I start defining *earthquake-induced migration* as the migration that occurs during the first 24 months after an earthquake. Later, I restrict my migration window to fourteen to six months after an earthquake.

1.4.3 Are Marriage Effects Different between Earthquake-induced Migrants and Left-behind?

Among the women exposed to earthquakes in my sample, 23% of them are *earthquake-induced migrants*. This figure reaches 37% when restricting the sample to women exposed to the highest earthquake ground shaking. Intuitively, we may expect to see different earthquake impacts between *earthquake-induced migrant* and *non-migrant*. A migration induced by external factors leads to a negative shock on households' income (see Table A.1.51). Furthermore, migration implies a new marriage market at the destination, where migrants' households lack local networks in the new market.

I study whether an earthquake affects differentially the marriage decisions of *earthquake-induced migrant* and *non-migrants* women. Ideally, we would need to compare migrant to non-migrant women both exposed to earthquakes and with similar pre-earthquake characteristics. Empirically, I compare *earthquake-induced migrant* to *non-migrant* women. To overcome potential differences between *earthquake-induced migrant* and *non-migrants* women, I control as many observable differences

[19] Only 0.69% of my sample crosses the Indonesian border after an earthquake.

as I can.

$$Y_{i,s,k,a} = \beta_0 + \beta_1 Eq_{s,p,a,m} + \beta_2 Eq_{s,p,a,m} * Disp_{i,s,a} + \beta_3 Eq_{s,p,a,m} * X_i + \alpha_p + \gamma_a + \delta_k + \zeta_u + \epsilon_{i,d} \tag{1.4}$$

The migration variable, $Disp_{i,s,a}$, is a time-varying measure of migration induced by an earthquake. $Disp_{i,s,a}$ is a dummy indicator that switches to 1 if individual i migrating induced for an earthquake occurring in sub-district of resident s and at a given age a, 0 otherwise. I control for urban fixed-effects at origin, ζ_{u_o}, to account for the different probabilities of marriage if your sub-district of origin is an urban location. And, I control for mother education and religion measured a year before an earthquake strikes and urban/rural destination, X_i.

While the baseline specification controls a bunch of observable characteristics, there may be unobservable differences between *earthquake-induced migrant* and *non-migrants* women that I cannot control for. To sharpen identification, I profit from the fact that some families have two or more daughters to include family-fixed effects. By designing a female siblings comparison age-by-age, I control for unobservable characteristics that otherwise wouldn't be feasible.[20] Appendix Table A.1.2 presents the descriptive statistics on household composition and characteristics. Notably, this approach allows me to account for regional characteristics at birth and residence at the moment of the earthquake, family attributes, household size, preferences, and networking capital. The within-family design also accounts for religion, ethnicity, culture, and social practices, which strongly correlate with marriage decisions in Indonesia.

1.5 Results

I present three sets of findings. First, I show the increase in annual child marriage hazards among young women exposed to earthquakes. Second, I give evidence on how earthquakes raise the probability of migrating. Third I describe how *earthquake-induced migrant women* are more affected by earthquakes compared to *non-migrant* young women.

1.5.1 The Effects of Earthquakes on Timing of Marriage

Table 1.1 reports the adverse effect of earthquakes on the annual marriage hazard for young women aged 12 to 22.[21] In column 3, I report the estimated coefficients

[20] Image a family with two daughters. In 2012, the entire family was exposed to an earthquake and migrated as a result. Daughter one was born in 1989 and married already in 2012 (at the age of 22). Daughter two was born in 1996 and non-married in 2012.

[21] By the age of 23, 79% of women are already married. Results do not change for a sample until age 24 (Table A.1.38)

for equation (1). It shows that women who experience an earthquake between ages 12 and 22 are 0.7 percentage points (pp) more likely to get married in the same year.[22] The effect is statistically significant at the 5% level. The average annual marriage hazard for this age group is equal to 0.036. Hence, the effect corresponds to an approximately 19% increase in the annual marriage hazard in response to an earthquake. In Figure A.1.3, I explore the heterogeneity of this effect by the woman's age by interacting earthquake with each age dummy. The strongest effects are at ages 16, 21 and 22. This is unsurprising because before 2019 the legal age at marriage for women in Indonesia was 16.

Table 1.1: Effect of earthquakes on the timing of marriage

	Below age 23			Below age 18		
	(1)	(2)	(3)	(4)	(5)	(6)
Earthquake	0.010***	0.008**	0.007**	0.010***	0.009***	0.008**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Dep. var. mean	0.036	0.036	0.036	0.018	0.018	0.018
Observations	585,816	585,816	585,816	350,232	350,232	350,232
Number of provinces	15	15	15	15	15	15
Number of years	22	22	22	22	22	22
Number of districts	255	255	255	255	255	255
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Urban FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth Year FE	No	Yes	Yes	No	Yes	Yes
Age FE	No	Yes	Yes	No	Yes	Yes
Controls	No	No	Yes	No	No	Yes

Note: This Table presents the earthquake results on the dependent variable: annual marriage hazard. The dependent variable is a binary variable for marriage, coded to one if the woman married at the age corresponding to the observation. Earthquakes are defined as earthquakes with an intensity of at least VII in some of its locations affected (Gignoux and Menéndez 2016). Observations are at the level of the person's age at the month level (from 12 to 22 or the age of first marriage). The baseline specification is presented in Equation 1. Column (1) presents the results with the province and urban fixed effects but without age, birth year fixed effects and covariates. Column (2) includes also age and birth year fixed effects. Column (3) controls for baseline characteristics (religion and mother education for the year before the earthquake). Columns (4), (5) and (6) perform the same analysis that Columns (1), (2) and (3) but for a sub-sample of ages from 12 to 17 (or the age of first marriage). Standard errors are clustered at the district level. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

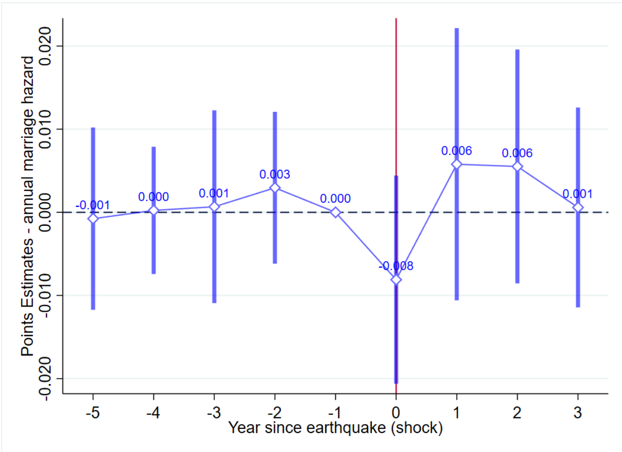
In column 6, I focus on child marriage (i.e., marriage below age 18). I restrict the panel dimension of my dataset to the ages between 12 and 17 and find that earthquakes have similar effects on the hazard of marriage at these early ages. Girls who experience an earthquake between the ages of 12 and 17 are 0.8 percentage points (pp) more likely to get married in the same year. The average annual marriage hazard for this age group is equal to 0.018. Thus, the effect corresponds to a 44% increase in the annual child marriage hazard.

The identification assumption is that absent treatment, the hazard of getting married for young women exposed to earthquakes would have evolved similarly to that of non-exposed women. I provide evidence supporting this assumption by estimating an event study version of the baseline specification that allows the effect to vary over time. Figure 1.2 reports β_p coefficient estimates from equation (2). The

[22] Equation (1) includes age-fixed effects in order to perform an age-by-age comparison.

figure shows no difference between exposed and non-exposed young women before an earthquake. There is a sharp decrease in the year of the earthquake, followed by a 0.6 pp increase in the annual marriage hazard in the first and second years after the earthquake. In the event study estimation, only exposed to earthquakes women are included. The differences in the sample could be the main reason for insignificant post-earthquake estimates. The baseline specification is also estimated using Chaisemartin and D'Haultfœuille 2020 estimators for accounting for the heterogeneous treatment.[23]

Figure 1.2: Effect of earthquakes on the timing of marriage, by year since treatment



Note: This figure plots the event and year coefficient from estimating equation 2 using timing of marriage as dependent variable. The confidence intervals are the 95%. Marriage outcomes comes from the IFLS and earthquake variation from USGS. The omitted category is T-1, earthquake year. The dataset is a person-age panel format. Treatment is defined at year level. Figure a present the estimates at the baseline specification.

Threats to Identification. There are at least two additional key identifying assumptions. First, the geographic location of an earthquake epicenter is plausible exogenous conditional on a certain location. We could be concern on potential discrepancies in the likelihood of earthquake occurrence within Indonesia provinces. To address this threat, I show that results hold when restricting my sample to locations with the highest seismic activity (see Appendix Table A.1.13). Second, conditional on the control variables, the difference-in-differences pick up the effect of an earthquake. Appendix Table A.1.3 shows that women exposed to earthquakes are older, better educated, have smaller families, and have better educated parents the year before an earthquake. However, they have lower household income and are less likely to own real estate (houses, farm, land). In my baseline specification I control for urban residence, age, year of birth, and mother's education. Adding these covariates allows me to control for some of these differences. In the robustness section, I run additional tests.

[23] Estimates are very similar for the Callaway and PedroSant'Anna 2021 and Sun and Abraham 2021.

A potential threat to the identification strategy comes from the treatment definition. In the baseline specification, the earthquake variable in equation (1), $Eq_{s,t}$, switches to 1 from the occurrence of an earthquake. Panel A of Appendix Table A.1.9 reports the results restricting the exposure to earthquakes from 0 to 11 years from the outset of an earthquake. The effects start from the second year and persist for 11 years thereafter.

Whereas the baseline specification focuses on the occurrence of earthquakes between ages 12 and 22, panel A of Appendix Table A.1.10 extends the exposure to earthquakes before the age of 12. The effects do not change. On average, women in the sample suffer only one earthquake between the ages of 12 to 22. However, 2.2% of women in my sample are exposed to more than one. Panel B shows the estimates for a continuous definition of earthquake exposure. The results remain unchanged.

There may be potential measurement error in reporting past marriages. I construct an alternative outcome using cohabitation data from the IFLS. The results are in line with the main estimates (Table A.1.11). In Indonesia, marriage migration is not a prevailing practice. Only 5.97% of women in my sample had migrated outside their village at the time of marriage. While marriage migration does not appear to be a major threat in Indonesia, another potential concern for the identification is whether women decide to marry before an earthquake but are obligated to delay their marriage after the disaster. Ideally, I should use data on when women make this decision. However, this data is not available. The IFLS gathers information on arranged marriage, and only 7.36% of marriages are arranged by the family. I use arranged marriage as a proxy for marriage decisions before an earthquake. In Table A.1.12, I remove arranged marriages from the sample. The estimates do not change.

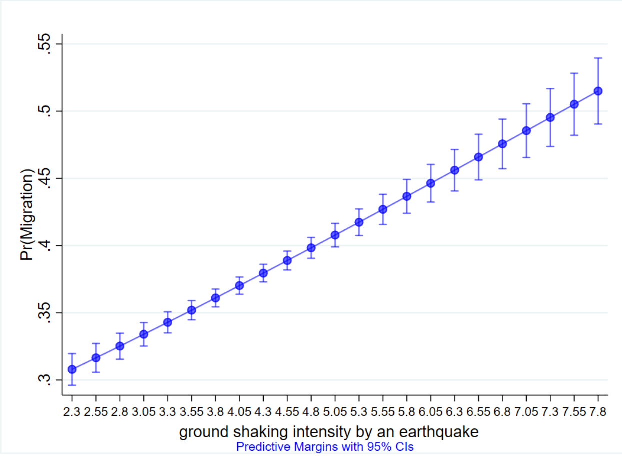
1.5.2 The Effects of Earthquakes on Migration Decisions

Table 1.2 shows the results from equation (3). I find earthquakes increase migration by 5.9 pp (significant at the 1% level, column 1). The effects are sizable in magnitude; the effects correspond to a 49% increase compared to the mean at baseline. This result is consistent with the literature on migration induced by natural disasters (Deryugina, Kawano, and Levitt 2018; Nakamura, Sigurdsson, and Steinsson 2022). Figure 1.3 illustrates that the probability of a household migration depends on the earthquake ground shaking and distance to the earthquake epicentre. The probability of migrating increases with the ground shaking (a), but decreases with the distance to the epicentre (b).

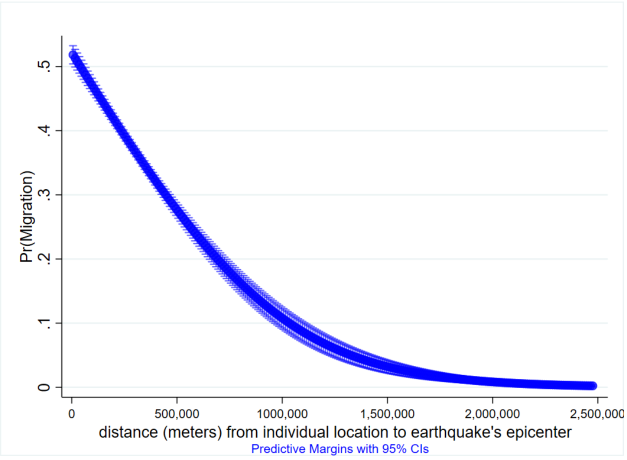
In columns 2, 3 and 4 of Table 1.2, I examine the heterogeneity of this effect by gender, ethnicity and religion by interacting the earthquake variable with a dummy.

Figure 1.3: Probability of migrating after an earthquake

a) By earthquake ground shaking



b) By distance to earthquake epicenter



Note: This figure presents the estimates on the probability of migrating after an earthquake (equation 3). Figure a shows the positive relationship between the probability of migrating and earthquake ground shaking at each household location. Ground shaking is measured using the Modified Mercalli intensity (from USGS). The Modified Mercalli intensity ranges from 0 to 10. Figure b shows the negative relationship between the probability of migrating and the distance to an earthquake epicentre. Both analyses include district fixed-effects. Results are unchanged when I redefine my migration timing window (from 6 to 24 months after an earthquake).

The results show that earthquakes do not induce a gender, ethnic or religion-driven migration.

While marriage migration does not appear to be a major threat in this context, another potential concern for my identification strategy is whether marriage migration grows during an earthquake. I find, in column 5, that earthquakes decrease marriage migration slightly (significant at the 1% level). In column 6, I present a further test. I study how age affects migration for women. Namely, I show how the number of years to turn age 23 affects the migration decisions for my main sample of women. Each additional year closer to age 23 decreases the decision to migrate by 0.6 pp (significant at the 1% level). This finding supports the hypothesis that women do not migrate to improve their marriage outcomes.

Table 1.2: Effect of earthquakes on migration decisions

	migration				marriage migration	migration
	(1)	(2)	(3)	(4)	(5)	(6)
Earthquake	0.059***	0.054***	0.020	0.055***	-0.019***	0.145***
	(0.020)	(0.020)	(0.025)	(0.021)	(0.003)	(0.025)
*Earthquake * Women*		-0.002				
		(0.005)				
*Earthquake * Non-Javanes*			0.064			
			(0.043)			
*Earthquake * Non-Muslim*				-0.027		
				(0.053)		
*Earthquake * years to 23*						-0.006***
						(0.001)
Dep. var. mean (1993)	0.120	0.120	0.120	0.120	0.120	0.120
Observations	162,600	162,600	162,600	162,600	162,600	33,415
Number of provinces	15	15	15	15	15	15
Number of survey years	5	5	5	5	5	5
Number of districts	255	255	255	255	255	255
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Year Survey FE	Yes	Yes	Yes	Yes	Yes	Yes
Urban FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Note: This Table presents the estimates from Equation 3. The dependent variable is a binary variable for migration, coded to one if the individuals move from their place of residence. Earthquakes are defined as earthquakes with an intensity of at least VII in some of its locations affected (Gignoux and Menéndez 2016). Observations are at the survey year level. Column (1) presents the main results. Columns (2), (3) and (4) report the heterogeneity results by gender, ethnicity, and religion. In Indonesia, 43% is Javanese and 87% Muslim (Population Census, 2010). Column (5) presents the results for migration as a consequence of marriages (*marriage migration*). Column (6) shows the results for women below 23 on an interaction with the age gap of 23. Standard errors are clustered at the district level. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

I perform the same analysis for a refined migration definition, *earthquake-induced migration*.[24] Appendix Table A.1.4 shows similar results. Moreover, results in Appendix Table A.1.21 are robust across different displacement definition windows.

1.5.3 Timing of Marriage: Induced Migrants versus Left-Behind

In this section, I explore if an earthquake differentially affects the marriage decisions of *earthquake-induced migrant women* and *left-behind* women. In section 5.1, women exposed to earthquakes include *earthquake-induced migrants* and *left-behind*. If I remove *induced-migrant* women from the sample, I find that the effects are concentrated below the age of 18. This approach enables me to show the net effects of earthquakes for those who stay in earthquake-affected areas. Girls who experience an earthquake between the ages of 12 and 17 are one percentage point (pp) more likely to get married in the same year. The effect corresponds to a 56% increase in the annual hazard of child marriage.

In panel A of Table 1.3, I present the results (for equation (4)) on how induced-migration bring young women's marriage foreward. I find that women who migrate to respond to earthquakes between the ages of 12 and 22 are 3.1 percentage points (pp) more likely to marry in the same year (column 1). Even after controlling for fixed effects by a cohort of birth, age (column 2) and covariates (column 3) (e.g.

[24] Namely, migration during the 24 moths after an earthquake.

being Muslim and mother's education a year before an earthquake strikes), I find that the effect of migration remains unchanged. The effects are significant at the 1% level and sizable in economic magnitude. The average annual marriage hazard for this age group is equal to 0.036, and the effect corresponds to an approximately 72% increase in the annual marriage hazard (column 3). The main counterfactual for the heterogeneity analysis is women that suffer an earthquake at their sub-district of residence at origin, but do not migrate. See Appendix Figure A.1.4 for an example. If I focus on child marriage, no statistically significant heterogeneous results exist between *earthquake-induced migrant women* and *left-behind women* (columns 4-6). In Figure A.1.3, I explore the heterogeneity of this effect by woman's age with the addition of an age dummy interaction. There is a jump at the age of 17. The lack of heterogeneous results does not mean that earthquakes do not affect annual child marriage hazard for *earthquake-induced migrants*. But, it means that the results on the probability of getting marriage below age, is not different from *left-behind* women.

Table 1.3: *Earthquake-induced migrants* versus *left-behind*

	Below age 23			Below age 18		
	(1)	(2)	(3)	(4)	(5)	(6)
			PANEL A: Main counterfactual			
Earthquake	0.005	0.002	-0.015**	0.012***	0.010***	0.009
	(0.003)	(0.003)	(0.006)	(0.004)	(0.004)	(0.008)
*Earthquake * Migration*	0.026***	0.026***	0.027***	-0.007	-0.008	-0.008
	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)
Observations	585,816	585,816	585,816	350,232	350,232	350,232
Birth Year FE	No	Yes	Yes	No	Yes	Yes
			PANEL B: Female siblings counterfactual			
Earthquake	0.031***	0.008	-0.028***	0.021***	0.007***	0.009
	(0.004)	(0.004)	(0.007)	(0.003)	(0.003)	(0.007)
*Earthquake * Migration*	0.043***	0.039***	0.038***	-0.004	-0.005	-0.004
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.010)
Observations	584,700	584,700	584,700	169,176	169,176	169,176
Family FE	Yes	Yes	Yes	Yes	Yes	Yes
Dep. var. mean	0.036	0.036	0.036	0.018	0.018	0.018
Number of provinces	15	15	15	15	15	15
Number of years	22	22	22	22	22	22
Number of districts	255	255	255	255	255	255
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Urban FE	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	No	Yes	Yes	No	Yes	Yes
Controls	No	No	Yes	No	No	Yes

Note: This Table presents the estimates from Equation 4 where *Earthquake * Migration* is the interaction of earthquakes with a migration right after an earthquake. The dependent variable is a binary variable for marriage, coded to one if the woman married at the age corresponding to the observation. Earthquakes are defined as earthquakes with an intensity of at least VII in some of its locations affected (Gignoux and Menéndez 2016). Therefore, the counterfactual is left-behind women. Observations are at the level of the person's age at the month level (from 12 to 22 or the age of first marriage). Panel A reports the results for the main counterfactual: left-behind women exposed to earthquakes. Column (1) presents the results with the province and urban fixed effects but without age, birth year fixed effects and covariates. Column (2) includes also age and birth year fixed effects. Column (3) controls for baseline characteristics (religion and mother education for the year before the earthquake). Panel B presents the results for girl-to-girl comparison within the same family. Columns (4), (5) and (6) perform the same analysis that Columns (1), (2) and (3) but for a sub-sample of ages from 12 to 17 (or the age of first marriage). Standard errors are clustered at the district level. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Threats to Identification. As I mentioned in section 4.2, the initial cutoff for the definition of earthquake-induced migration is 24 months after the earthquake. Many other events could happen in between, however the data does not allow me to control for these changes. I am aware that the selection of the migration cutoff is arbitrary. To overcome this limitation, I repeat the analysis for a set of cutoffs. The results hold when I restrict the migration window from 6 to 14 months after an earthquake (Appendix Table A.1.20).

Although restricting the earthquake sample to those with large enough strength reduces the scope of self-selection, an obvious concern is that other factors besides earthquakes may drive the decision to migrate.[25] In order to overcome this caveat, I perform a within-family analysis comparing women with their female siblings (more details in section 4.3). Panel B of Table 1.3 presents the results. Women who migrate induced by an earthquake between the ages of 12 and 22 are 3.9 percentage points (pp) more likely to get married in the same year (column 2). The effects are significant at the 1% level and correspond to approximately a 108% increase in the annual marriage hazard. In line with the baseline results, no statistically significant heterogeneous findings exists between *earthquake-induced migrant women* and *left-behind women*, when restricting the panel dimension to child marriage (columns 5). In Appendix Table A.1.24, I conduct further sanity checks exploiting the age gap between siblings, and number of female siblings. The net effects when including age gap interaction (column 1) or restricting to a sub-sample of families with two daughters (column 5) are considerably larger. Closer-in-age daughters or a higher number of daughters creates a substitution effect between siblings. Hence, when the age distance increases and the number of daughters decreases to the minimum the potential of marriage strategies to cope with migration shock decreases.

A certain level of selection *into destinations* by *earthquake-induced migrant women* could potential exist. As Appendix Table A.1.5 shows, the majority (68.66%) of *earthquake-induced migrant women* in my sample settled within the same district, and about one-third moved outside their district of origin (i.e. within the same province, to other IFLS provinces or other non-IFLS provinces). It would be a limitation if the young women less prone to getting married moved to other provinces not in the sample, leading to biased results. To address this concern, I begin by noting that only 3.39% of *earthquake-induced migrant* women move to a non-IFLS province. Appendix Table A.1.25 reports the estimates of equation (4) by a sub-

[25] Appendix Table A.1.22 presents differences between *earthquake-induced migrant women* and *left-behind* women. The *earthquake-induced migrant women* are, on average less educated, younger, less likely to self-report as Muslim and more likely to have higher savings before an earthquake. However, conditional to being equally exposed to an earthquake, *earthquake-induced migrants* have less financial capacity, more likely to be self-employed working in the agricultural sector, but are better educated (see Table A.1.23).

sample of destinations (column 1, within the same district; column 2, within the same province; column 3, to another IFLS province). The estimates are positive and significant, except for column 3. In columns 4 and 5, I include two different interactions. First, I interact with the district area for those that move within the same district and, second, with distance to destination for those that do not move within the same district. I find that district area and distance to destination do not affect the results. Overall, these results suggest that the selection of displaced women into different destinations does not drive the results.

1.5.3.1 Placed-Based Effects on Marriage Decisions

To provide further checks on how the destinations might affect my results, I quantify the relative importance of the local marriage market to the marriage decisions of displaced women. To capture place-based effects, I exploit origin-destination differences in development (population density, night light intensity), marriage market composition (ethnicity composition, sex ratio, unmarried population), and earthquake intensity (ground shaking and houses destroyed). To estimate the effects, I interact $Eq_{s,p,a,m} * Disp_{i,s,a}$ to destination-origin differences, Δ_{od}, in equation 4. Panel A, B, and C of Appendix Table A.1.26 shows the results.

Some 72% of women move within their ethnic homeland or to another homeland adjacent to theirs. Panel A shows that the net effects do not change if the destination falls within their homeland (column 1), within or adjacent to their homeland (column 2), whether the origin is in their homeland but the new destination is not (column 3), or where the origin and destination are both in their homeland (column 4). However, estimates are 1.5 times larger when the destination is in their homeland but the origin is not (column 5). A possible interpretation of these results is the loss of social capital within their own ethnic group during their stay outside their ethnic homeland. Therefore, women pull forward their marriage in order to integrate again.

In Panel B, I show that the economic development at the destination slightly increases the effects of displacement. There is an increase of around 19% when studying differences in population density between origin and destination after an earthquake (column 2), differences in population density between origin before an earthquake and destination after an earthquake (column 3) and differences in night light intensity between origin and destination after an earthquake (column 5). When I only control population density (column 1) or night light intensity at the destination after an earthquake (column 4), the increase corresponds to 53%.

In Panel C, I evaluate how the marriage market composition at the destination could affect the results. Columns 1, 2 and 3 show the sex ratio at the destination after an earthquake (column 1), differences between origin and destination after an earthquake (column 2), and differences between origin before an earthquake and destination after an earthquake (column 3) do not change the results. However, the fact that an earthquake also hits the marriage market at the destination slightly increases the results (column 4).[26]

1.6 Mechanisms

In this section, I present the main mechanisms underlying the results. While *earthquake-induced migrant women* and *left-behind women* are both affected by earthquakes, the income shock of migrating exacerbates the poor economic outcomes of earthquake-affected women. See Appendix Table A.1.51 for more details (Nakamura, Sigurdsson, and Steinsson 2022; Deryugina, Kawano, and Levitt 2018). On top of that, migrant women end up in a new marriage market in which they have no local networks to turn to. To explain the heterogeneous earthquake's effects on the timing of marriage, I present three forms of informal insurance that *earthquake-induced migrant women* can obtain from their marriage: marriage transfer, an increase in aggregate labour return when newly formed couples join the bride's household at marriage, and integration with the local population at destination (sections 6.1, 6.2 and 6.3). They suggest that marriage is a strategy used to cope with their migration shock. In Appendix Section F.1.1, I develop a simple model reconciling the empirical results. Appendix Section F.2 rules out other mechanisms. Furthermore, Section 6.4 describe the main mechanisms for left-behind women.

1.6.1 Bride Price to Alleviate Financial Constraints

Traditional marriage payment norms determine women's age at marriage when families face adverse shocks (Corno, Hildebrandt, and Voena 2020). In Indonesia, the tradition exists that the bride's family receive a transfer from the groom at marriage (called the bride price). However, bride price payments may be endogenous to the economic circumstances at the time of marriage, notably migration. To test how bride price payments change the effects of displacement, I follow Ashraf et al. 2020, who exploit historical data on heterogeneity in marriage payments across ethnic groups from the Ethnographic Atlas (1967). I circumvent a fundamental empirical challenge by using historical information on marriage payments at the ethnicity level rather than actual payments.

Panel A in Table 1.4 presents the estimated heterogeneous effects of earthquakes on the timing of marriage between *earthquake-induced migrant women* and *left-behind women* adding another dimension of heterogeneity: the prevalence of the

[26] I use the Population Census from 1990, 2000 and 2010 to calculate the sex ratio for the unmarried population below the age of 23.

traditional bride price custom within Indonesia. Column 2 shows the effect for a sub-sample of women traditionally engaged in the bride price custom. I find that the bride price tradition matters: the effect is much stronger for *earthquake-induced migrant women* from groups that are traditionally engaged in the bride price. Strikingly, the annual hazard into marriage is 44% higher among the bride price sub-sample compared with the non-bride price sub-sample. Hence, the results show that the aggregate effect masks substantial heterogeneity that depends on a group's marriage customs. When restricting the panel dimension to ages below 18, the point estimates are non-significant.[27]

Table 1.4: Effects on the timing of marriage, by marriage norms

	(1)	(2)	(3)
	PANEL A: Bride Price tradition		
	All sample	Bride Price	Non-Bride Price
Earthquake	-0.015**	-0.019**	0.003
	(0.006)	(0.009)	(0.020)
*Earthquake * Migration*	0.027***	0.037***	0.025***
	(0.005)	(0.008)	(0.006)
Observations	585,816	120,698	455,583
	PANEL B: Matrilocality tradition		
	All sample	Matrilocal	Non-Matrilocal
Earthquake	-0.015**	0.099	-0.024***
	(0.006)	(0.252)	(0.008)
*Earthquake * Migration*	0.027***	0.165	0.027***
	(0.005)	(0.127)	(0.006)
Observations	585,816	4,055	500,469
Dep. var. mean	0.036	0.036	0.036
Number of provinces	15	15	15
Number of years	22	22	22
Number of districts	255	255	255
Province FE	Yes	Yes	Yes
Urban FE	Yes	Yes	Yes
Birth Year FE	No	Yes	Yes
Age FE	No	Yes	Yes
Controls	No	No	Yes

Note: This Table presents the estimates from Equation 4 by marriage norms: *bride price* and *matrilocality* traditions. *Bride price* tradition is a payment from the groom (or groom's family) to the bride (or bride's family) at the moment of the marriage. In Indonesia doesn't exist a payment from the bride to the groom's family (*dowry*). *Matrilocality* tradition is whereby the husband joins the wife's household after the marriage. When the wife joins the husband's household or settles down in a new household is known *patrilocality* or *neolocality*. The dependent variable is a binary variable for marriage, coded to one if the woman married at the age corresponding to the observation. Estimates include province, urban, age and birth year fixed effects and control for baseline characteristics (religion and mother education for the year before the earthquake). Therefore, the counterfactual is left-behind women. Observations are at the level of the person's age at the month level (from 12 to 22 or the age of first marriage). Panel A reports the results by *bride price* sub-sample. Panel B reports the results by *matrilocal* women sub-sample. Standard errors are clustered at the district level. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Estimates on marriages below the age of 18 are non-significant.

One potential explanation of the findings could be the consumption smoothing mechanism that marriage payments mean for *earthquake-induced migrant house-*

[27] Appendix Table A.1.28 reports the results with a bride price interaction. Estimates are non-significant.

holds. Results of section 5 highlight that the heterogeneous effects of earthquakes start at age 18. Why do the effects start when women turn 18? Do payments at marriage that *earthquake-induced migrant* women receive change across the life cycle? Data from the Indonesia Family Life Survey (IFLS) provides information on bride price payments for each couple's marriage. I find that both bride price and non–bride price groups tend to report positive payments at marriage but that there are noticeable differences in the size of payments between the two groups. Following Corno, Hildebrandt, and Voena 2020, I use the natural logarithm of the marriage payment for bride price women as the dependent variable in the equation (4).

In the IFLS data, the mean bride price for the sample of 8,608 women equals 7,597,882 Indonesian rupees, with a standard deviation of 3.25.[28] The payment received by bride migrant households compared to *left-behind women* do not change when estimating the effects for the entire sample (estimates are statistically non-significant). Nevertheless, payments increase 55% when restricted to a sub-sample of women traditionally engaged in the bride price custom (Appendix Table A.1.6, Column (4)). Migrant women receive a higher payment with respect to *left-behind women* because earthquake-affected areas are credit constrained after the disaster. However, the net effects in column 6 are negative. The net changes in prices experience a 38% statistically significant decrease with respect to the price of *left-behind* women. Adding controls for religion and mother's education before the earthquake (column 5) and current education fixed effects (column 6) substantially change the estimates. The results suggest that women's education is a key component of the returns from marriage (Goldin 2006; Chiappori, Salanié, and Weiss 2017; Ashraf et al. 2020).[29] Estimates are positive and non-significant for marriages below the age of 18.

Panel C of Appendix Table A.1.6 shows the compositional effects by groom (*earthquake-induced migrant* or *native*). Migrant grooms drive this increase. A possible reason for these results is the trade-off between getting married and higher prices for migrant grooms. *Earthquake-induced migrant grooms* may not be competitive in the market with the *native* bride. Moreover, women can gain better potential networks by marrying a *native*. Therefore, the migrant groom may be willing to increase the payment if they want to marry at the new destination.[30] This finding aligns with the model's predictions in Section F.1.1.

Are marriage payments that migrant women receive indeed higher than those for *natives*? To answer this question, I design a new counterfactual: non-exposed to earthquakes *native* women. Appendix Table A.1.32 report the results. Results are

[28] I calculate real prices with the baseline year 2000.

[29] Appendix Figure A.1.7 shows that education is a key factor for bride price and matrilocal women.

[30] Appendix Figure A.1.3 shows the heterogeneity results by women's age.

significant when including current education fixed effects (column 6). There is a 26% increase in the price compared to *native* women. For this exercise, the increase gives the impression of being driven by *native* grooms (Panel C). The intuition of these results is that a disaster-induced migration drops income for migrants grooms. And, *earthquake-induced migrant* women increase the minimum amount they are willing to accept to marry earlier. An increase in the transfer they receive help household alleviate their financial constraints.[31] Unfortunately, I cannot disentangle why the payment at marriage is higher among displaced versus *native* women when I control women's education. A potential channel may be the exotic aspect of the newcomers. However, I cannot test it empirically.

1.6.2 Matrilocality: An Increase in Household Labour Return

In the matrilocality tradition, newly formed couples join the wife's household after marriage.[32] Hence, marrying women earlier can lead to an additional productive household member in the family economy (i.e., the new son-in-law). The marginal increase in household labour return from the son-in-law can help the *earthquake-induced migrant* household cope with their migration income shock. Could the coping channel from the matrilocal tradition drive my results?

I analyse whether the post-marriage residency tradition could affect the results. I test if differential impacts exist on the timing of marriage between women traditionally engaged in matrilocal and non-matrilocal customs. To do so, I use historical data on the residency traditions of ethnic groups from from the Ethnographic Atlas (1967). Panel B in Table 1.4 presents the results. I find that *earthquake-induced migrant* women from ethnic groups where the groom resides with the bride's parents are more responsive to earthquakes. The effects are seven times larger among matrilocal women compared with patrilocal and neolocal women.[33] Results are non-statistically significant for a matrilocal sub-sample. Only 5% of the sample is traditionally engaged in matrilocality, so I lose a lot of power. However, Appendix Table A.1.28 reports the results with a matrilocality interaction. The estimates are positive and significant.[34]

Why do the heterogeneous effects between *earthquake-induced migrant* and left-behind women start at the age of 18? Do marriage returns from matrilocal residence

[31] Results are non-significant for marriages below age 18. Panel B presents the results.

[32] Matrilocal residence denotes a tradition whereby the husband joins the wife's household after marriage. When a wife join the husband's household this is patrilocal residence. Neolocal residence is when husband and wife reside apart.

[33] From an anthropologist's perspective, there is a strong relationship between matrilocality and bride price traditions. I perform the same analysis for a sub-sample of bride-price women. The results do not change (Appendix Table A.1.29).

[34] Appendix Figure A.1.6 shows that the heterogeneous effects for *earthquake-induced migrant* women traditionally engaged in matrilocality start from the age of 18.

change across the life cycle? Education increases the labour return. Therefore, I use the groom and bride's education at marriage to assess if there are differences in the education-matching between the spouses at marriage. The hypothesis is that matrilocal *earthquake-induced migrant women* marry better-educated men to increase the groom's labour contribution to their household. Appendix Table A.1.7 finds that controlling for women's education increases the groom's education at marriage for matrilocal women who migrate after the disaster (columns 2 and 5). Estimates are statistically significant at the 1 and 5% levels. But, the net effect is zero when women are below age 23 and becomes negative if below age 17. These results suggest that matrilocal migrant women marry higher educated men, and their marriage return turns higher when women reach adulthood.[35]

1.6.3 Integration with the Local Population

Earthquake-induced migrant women lack local networks that could help them after the shock. Therefore, expanding their extended family to new members at their destination may function as informal insurance and allow them to find a job at the destination. The anticipation of women's marriage can be a quick way to integrate with the local population at the destination.

To study this potential channel, I proxy local engagement with data on involvement in a community organisation from the IFLS. I use data on parents' participation in an *arisan* in the last 12 months, the number of arisan, and participation in community groups in the last 12 months.[36] I add an additional interaction in equation (4) between earthquake, induced migration and involvement in a community organisation. Table 1.5 reports the results. I find that the three interactions are negative, but it is only significant (at the 1% and 10% level) the interaction to the participation in community groups (columns 3 and 6). Although the net effects are marginally greater, the involvement in a community organisation decreases the effects of earthquakes by a 12% of the baseline results in section 5.3. This result suggests that being engaged at the destination mitigates the heterogeneous effects of earthquakes on the timing of marriage. Therefore, the finding is in line with the hypothesis that women marry earlier to build networks at the new destination. [37]

[35] Like with the bride price mechanisms, I compare *earthquake-induced migrant women* to never-exposed *native women*. The results report no effect on the spouse's education (Table A.1.33).

[36] An *arisan* is a social club, primarily populated by women. Members have similar backgrounds or interests. It represents an alternative to bank loans and other forms of credit.

[37] We might expect to see an increase in married couples from different ethnic groups as an alternative strategy to integrate. However, there is no evidence supporting this hypothesis (see results in Appendix Table A.1.31).

Table 1.5: Integration with local population, mitigation effects

	Below age 23			Below age 18		
	(1)	(2)	(3)	(4)	(5)	(6)
Earthquake	-0.021***	-0.020***	-0.021***	0.007	0.007	0.007
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
*Earthquake * Migration*	0.031***	0.031***	0.034***	-0.004	-0.004	0.001
	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)
*Earthquake * Migration * Arisan*	-0.009			-0.008		
	(0.010)			(0.007)		
*Earthquake * Migration * Nº arisan*		-0.002			-0.002	
		(0.002)			(0.001)	
*Earthquake * Migration * Nº com. act.*			-0.003*			-0.005***
			(0.002)			(0.002)
Observations	585,816	585,816	585,816	350,232	350,232	350,232
Dep. var. mean	0.036	0.036	0.036	0.018	0.018	0.018
Number of provinces	15	15	15	15	15	15
Number of years	22	22	22	22	22	22
Number of districts	255	255	255	255	255	255
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Urban FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Note: This Table tests the hypothesis that earthquake-induced migrant women may anticipate their marriage as a way to facilitate their family's quick integration with local populations at the destination. This Table presents the estimates from Equation 4 with an interaction to a variable on their family involvement in local communities at the new destination: *arisan*, *number of arisan*, and *number of community organizations* women's family participate in. An *arisan* is a social club that provides alternative bank loans and other forms of credit to its members. The dependent variable is a binary variable for marriage, coded to one if the woman married at the age corresponding to the observation. Therefore, the counterfactual is left-behind women. Observations are at the level of the person's age at the month level (from 12 to 22 or the age of first marriage). The estimates include province, urban, age and birth year fixed effects and controls for baseline characteristics (religion and mother education for the year before the earthquake). Column (1) presents the results with interaction equal to 1 if the women's family participates in an *arisan*. Column (2) includes an interaction to *number of arisan* that women's family participates in. Column (3) presents the results with interaction *number of community organizations* women's families participate in. Columns (4), (5) and (6) perform the same analysis that Columns (1), (2) and (3) but for a sub-sample of ages from 12 to 17 (or the age of first marriage). Standard errors are clustered at the district level. Robust standard errors are in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

1.6.4 Welfare Effects from an Early Marriage

The three mechanisms presented above show how *earthquake-induced migrant young women* marry earlier as a strategy for coping with the shock of their migration. Nonetheless, it is unclear if their early marriage directly translates into a positive welfare effect for them and their households.

The early marriage of *earthquake-induced migrant women* directly affects them and their families. By comparing married and unmarried migrant women from the same cohort, age and province, I show that married women have their first child earlier and have less likelihood of being employed. I also find that the consumption capacity of *disaster-induced migrant households* do not change with the early marriage of their daughter. If so, income from non-labour activities decreases, and labour income does not change. Food and non-food expenditure remain unchanged too. Migration also affects the characteristics of marriages. Migrant women are more likely to have lower education than their husbands and marry a fellow migrant compared to *left-behind women*. However, they are less likely to be in a polygamous marriage than women who are *left-behind* in earthquake-affected areas. These findings imply that married women and their households end up worse off compared to unmarried women. In Appendix Section F.4, I describe the analysis.

1.6.5 Left-behind Women

I find no evidence of the financial mechanisms above presented for *earthquake-induced migrant women* being present for *left behind women* (see Section F.3 for further details). The effects for *left-behind women* have to do with a response to a supply shock. I present suggestive evidence of two main mechanisms.

Population Outflow. Earthquakes trigger a large outflow of population, which changes the demographic composition of the marriage market. However, the effects of earthquakes on the migration decisions of individuals are not conditional on their gender, religion, or ethnicity. Although an earthquake does not affect the sex ratio, earthquakes increase women's fear of not finding a good match in the future. Hence, households may prefer to marry their daughter in childhood, rather than waiting and risking not finding a good match in the future. I use population data from the Indonesian Population Census (1990, 2000, and 2010) to measure district-level population changes. I employ two measures: changes in total population and the sex ratio of the unmarried population below age 23 before and after an earthquake. Appendix Table A.1.42 presents the results. I find that an increase in the population creates a net increase in the annual marriage hazard (Panel A). However, Panel B shows how the sex ratio does not affect the results. Unfortunately, due to data limitations, these results are just suggestive evidence and do not allow me to leverage conclusive evidence.

Schools Destruction. Destructive earthquakes affect a wide range of infrastructure and can potentially destroy public buildings, including schools. Due to the disruption in the supply of education this causes, school attendance decreases, and consequently, schooling drops. Girls are especially vulnerable to dropping schools in the aftermath of a disaster Takasaki 2017. I study how earthquakes affect school attendance and educational attainment. Appendix Table A.1.46 shows how earthquakes decrease school attendance in women below the age of 18 (Panel A). But, earthquakes do not affect educational attainment (Panel B). Girls who have already dropped out of school are more likely to marry. These findings suggest that a drop in schooling can explain an earthquake's effect on marriage below the age of 18 for women.

1.7 Can Household Incentives Mitigate the Effects on Early Marriage?

So far, I have examined how earthquakes anticipates women's marriage, as well as, its heterogeneous effects for forcibly displaced women. In this section, I study how policy can address the underlying mechanisms that lead *earthquake-induced migrant women* to marry earlier by changing household incentives. I exploit the differential timing of the implementation of an unconditional cash transfer (UCT) program from 2005 to 2014. This analysis helps further disentangle the mechanisms of the previous results (whether they are driven by an economic compensation) and sheds light on the role of policy to mitigate the effects of *earthquake-induced migrant women*.

1.7.1 Unconditional Cash Transfers

I show that providing monetary incentives for *earthquake-induced migrant households* mitigates the effect of earthquakes on early marriage. In particular, the UCT program reduces the annual marriage hazard. I also provide suggestive evidence that the UCT has larger mitigation effects among *earthquake-induced migrant households* engaged in the practice of bride price.[38]

The UCT program consists of a transfer to the poorest households of Rp 1.2 million for one year, provided on a quarterly basis (Rp300,000 per three month).[39] Figure 1.4 shows the time horizon of the IFLS data and cash transfer disbursements in Indonesia. The targeting process of the unconditional cash transfer program in 2005 used a Socioeconomic targeting tool.[40]

The identification challenge to analyse the effects of UCT program is the fact that the program was not randomly assigned. Hence, to circumvent the selection bias, I implement a similar strategy as in Bazzi, Sumarto, and Suryahadi 2015.I use the non-beneficiary households that have observably similar pre-program characteristics in 2000 as a control group. In order to balance the characteristics of the treatment and control groups, I then match the pre-UCT characteristics between the beneficiaries and non-beneficiaries based on the UCT beneficiary status from the IFLS's wave 4 (2007) and 5 (2014) using Coarsened Exact Matching (CEM).[41]

[38] Unconditional Cash Transfers do not affect migration decisions. The results can be provided upon request.

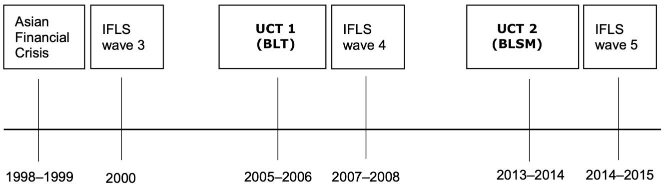
[39] That benefit is around 15% of the quarterly expenditures for the average beneficiary (Bazzi, Sumarto, and Suryahadi 2015).

[40] Socioeconomic targeting 2005 tool (Pendataan Sosial-Ekonomi) is a survey that was conducted by Statistics Indonesia (Badan Pusat Statistik or BPS). BPS collected 14 non-monetary variables to measure the welfare of poor households. A similar survey was also conducted in the next UCT programs.

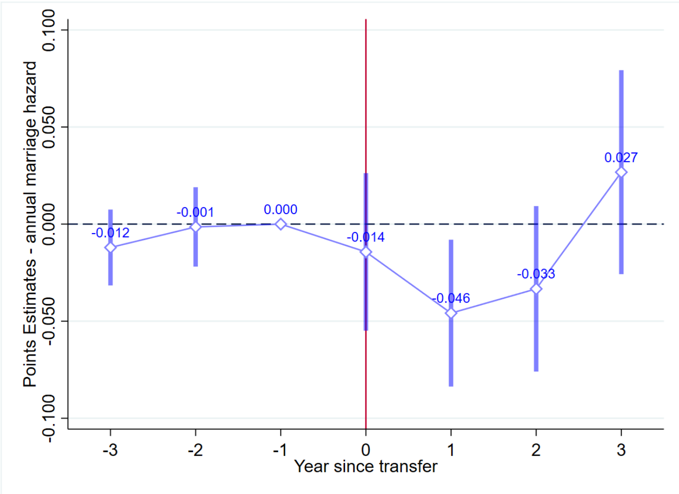
[41] I use the variables from Bazzi, Sumarto, and Suryahadi (2015) for the matching. These variables significantly affect the likelihood of households to receive the unconditional cash transfers.

Figure 1.4: Effects of Unconditional Cash Transfers (UCTs)

a) Time horizon of the IFLS data and UCT disbursements, Indonesia



b) UCT effects on the timing of marriage, *earthquake-induced migrant* women



Note: Figure 5 panel (a) presents the time variation in the Unconditional Cash Transfer disbursements in Indonesia, and the IFLS's waves. Panel b plots the event and year coefficient from estimating the effects of UCT in a sub-sample of *earthquake-induced migrant women* using annual marriage hazard as outcome variable. In compare UCT-beneficiaries and non-beneficiaries using Coarsened Exact Matching (CEM). I report coefficient estimates and 95% confidence intervals. The omitted category is T-1, uct first year. Standard errors are clustered at district level. The dataset is a person-age panel format. UCT disbursement is defined at year level.

I start by estimating the following empirical specification:

$$Y_{i,s,k,a} = \beta_0 + \beta_1 Eq_{s,p,a,m} + \beta_2 Eq_{s,p,a,m} * Disp_{i,s,a} + \beta_3 Eq_{s,p,a,m} * Disp_{i,s,a} * UCT_{ia} \tag{1.5}$$

$$+\beta_4 Eq_{s,p,a,m} * X_i + \alpha_p + \gamma_a + \delta_k + \zeta_u + \epsilon_{i,d}$$

where UCT_{ia} is a dummy indicating whether the individual i received a UCT at age a[42]. Column 1 of Table A.1.8 presents the results for the unmatching and CEM estimates. I find that UCT do not change the effects of earthquakes on the timing of marriage. I then proceed to evaluate the potential heterogeneity between *earthquake-induced migrant women* and *left-behind women* by sub-samples of UCT-beneficiaries and nonbeneficiaries with similar pre-program characteristics. Column 2 and 3 show how the results do not hold for UCT-beneficiaries sub-sample. This findings may indicate that UCT could help households to cope with their migration shock. As a next step, I run an event study analysis for a sub-sample of *earthquake-*

[42] I further restrict my sample for the UCT period (2005 to 2014)

induced migrant women. I examine whether there are some lag effect in the timing of marriage from the first year of the transfer. The results of Figure 1.4 suggest that UCT mitigates the anticipation effects of earthquakes on the timing of marriage of *earthquake-induced migrant* women. Next, I analyse whether this reduction differs by marriage traditions. Figure A.1.8 illustrates how the mitigation effects are larger among bride price women, but there are not differences among matrilocal and non-matrilocal women.

These findings have several implications. On the one hand, they help to disentangle the mechanisms behind the main estimates. If marriage effects were driven by other factors apart from the economic shock of moving, increasing the financial capacity of *earthquake-induced migrant households* would not mitigate the effects. One the other hand, these results shed light on the role of policy. Policy makers can potentially assist *earthquake-induced migrant women* by stimulating the development of formal financial coping strategies. Not only does this have implications for reducing women's early marriage. But these policies could also reduce the welfare consequences for women and their children.

1.8 Robustness Checks

Earthquake Exogeneity. A major identifying assumption is that the timing and the epicentre location of an earthquake in a specific province are assumed to be uncorrelated with other determinants of changes in the timing of marriage. In particular, we could be concern that households could choice their places of residence based on the likelihood of large earthquakes. Figure A.1.10 shows that earthquakes occur all over Indonesia, however, there is a high concentration on the west islands. To overcome this limitation, Table A.1.13 shows that the result of earthquakes on early marriage holds when restricting my sample to three different definitions of areas highly exposed to earthquakes: Sumatra island, the western island; provinces closer to seismic plaques; and districts closest to seismic plaques.

Earthquake Definition. I evaluate if my results hold with different definitions of destructive earthquakes. In Figure A.1.11, I change the ground shaking cut off that I use to define destructive earthquakes. I move the cut-off from 7 or more to 3 or more and iterate over a range. The statistically significance and magnitude of the effects persists along the five different definitions.

Inference. To account for the potential correlation in error terms across space between different geographical units, I consider clustering my standard errors at the province level and at the island level respectively. I also compute wild bootstrapped p-values following Cameron, Gelbach, and Miller 2008. In Table A.1.14, I replicate the estimates from Table 1.1 and report the corresponding p-values. The clustering exercise does not affect the statistical significance of my estimates.

To check whether my results hold when changing my specification, I run my baseline empirical specification from Table 1.1, changing my location fixed effects and clusters. Table A.1.15 demonstrates that the results are unchanged. In Table A.1.16, I perform a similar exercise changing the covariates in my main specification of equation (1). If I include father's education and number of siblings before an earthquake the effects on marriage below the age of 18 does not hold any more. There are two possible explanations for this. First, when including these covariates there is a big drop in the sample, and potentially affecting the estimation power. Second, father's education and number of siblings is strongly affected by mother's education with potential collider problems.

To study whether my findings hold with different sample definitions, I examine the analysis with a new sample of women. In the new sample I include women that are at least 25 years old at the last interview. I find that results do not change Table A.1.38 with the alternative sample.

A standard challenge in the literature studying migration is the definition of a counterfactual.[43] I estimate the effects of Table 1.3 restricting the counterfactual to later *earthquake-induced migrants* (Panel B of Table A.1.27) and non-exposed to earthquakes (Panel C). On the one hand, estimates in Panel B become insignificant. This paper conducts a within province analysis, and only a 2.2% of my sample is exposed to more than one earthquake. Therefore, the fact that the estimates turn insignificant is unsurprising. On the other hand, the points estimates are larger when non-exposed to earthquakes women function as the counterfactual. The change in the magnitude of earthquake effects captures two main differences between the counterfactual and *earthquake-induced migrant women*: the exposure to earthquakes, and their migration shock. Using non-exposed women as a counterfactual I can not disentangle if the effects are driven by one or another difference. Because the estimates capture both at the same time.

We may be concern that the effects of earthquakes could be affected by potential unobservables. I remove from the sample voluntary migrants, in Table A.1.18, and *earthquake-induced migrant women* who return to the sub-district of origin, in Table A.1.19. Results hold for both specifications.

Cultural Norms. I test for the possibility that bride price or matrilocality traditions might be correlated with other ethnicity-level characteristics that could lead to differential effects. In Table A.1.34, I evaluate the effects by whether an

[43] In the literature, the control group for displaced individuals can be stayers (e.g. Kondylis, 2010), residents in adjacent places to the affected areas (e.g. Fiala, 2015), with ex-ante no different than displaced individuals but non-affected by the shock (e.g., Sarvimäki, 2009; Bauer et al., 2013), or there is simply no control group at all (e.g., Ibañez and Moya, 2010)

ethnic group traditionally has a significant female participation in agriculture (matrilineality) or there is a polygyny tradition. A tradition of polygyny does not affect the results. Estimates are non-significant among matrilineality women.

I check the robustness of my estimates to the use of alternative measures for bride price and matrilocality. I construct new measures using contemporaneous data on bride price payments and change of household at marriage. The new bride price variable is 1 if a woman received a marriage payment from the groom (0 otherwise). The new matrilocal variable is 1 if a woman remains in her household after marriage (0 otherwise). The results of Table A.1.35 are only in line for matrilocality tradition. The results on bride price are inconsistent with the baseline results. A potential reason for this finding is the fact that both bride price and non–bride price groups tend to report positive payments at marriage but that there are noticeable differences in the size of payments between the two groups.

Placebo Groups. If indeed a marriage transfer from the groom and the tradition where husband joins wife's household matter for the effects of earthquakes among *earthquake-induced migrant women*, we should not see the same relationships in the data for men as we do for women. I test if this is true by replicating my analysis using a sample of men. I begin by ruling out the hypothesis that earthquakes impact the marriage market of men, to then, studying whether the heterogeneous results for bride price men hold.

I replicate the analysis reported in Table 1.1 using the sample of men.[44] As reported in Table A.1.40, I do not find the same patterns among men. Panel A shows no effects of earthquakes on the timing of marriage of men. And, there are similar heterogeneity effects between *earthquake-induced migrant men* and *left-behind men* compared to women (Panel B). In contrast to the case with women, for men we do not find a significant relationship for bride price men.

I run two additional placebo texts exploiting women's migration data and age variation at the moment of an earthquake. As expected, Panel A of Table A.1.39 shows no results of an earthquake exposure above the age of 22. By the age of 23, 81% of women are already married. However, Panel B presents a negative correlation between voluntary migration and annual marriage hazard. A planned ahead migration does not necessarily come with a drop in income and social networks. Hence, women's incentives to move ahead their marriage decrease.

1.9 Conclusion

This paper provides evidence that earthquakes raise women's early marriage. However, the overall effects mask substantial heterogeneity: the effects are higher for

[44] I keep all men that were at least 23 in the last round and were born after 1980.

earthquake-induced migrant women compared to *left-behind women*. I show that *earthquake-induced migrants* marry earlier as a financial coping strategy, gaining in the process: a marriage payment, an increase in labour return when the husband joins the household, and social integration in receiving communities. I also argue that policies like Unconditional Cash Transfers can mitigate the impact of destructive events on *disaster-induced migrant women*'s early marriage and, therefore, on their welfare.

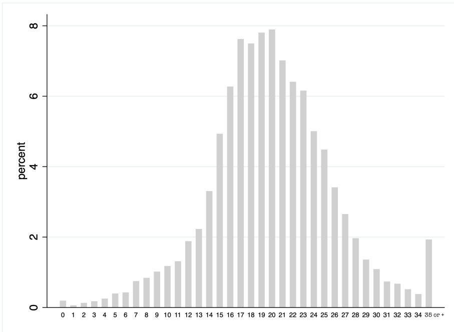
Though the situation in Indonesia is unique in many ways, the results of this paper can be applied to other countries. Entire populations are already suffering the impacts of climate change. But, many others will be affected in the near future. Unfortunately, the number of disasters is growing in frequency. On top of that, these results could be also applied, at a certain extend, to population movements driven by conflicts or human made disasters. Namely, because, there are many other examples of forced migration settings whose marriage markets are dominated by marriage customs that might have similar unintended consequences for young women. For instance, Ethiopian households are strongly affected by drought/floods and ethnic conflicts. Factors that are the leading causes of internal displacement in Ethiopia (Tesfaw 2022). A country where wedding customs vary among the diverse tribes of the country. Similarly, in Myanmar, where Rohingya Muslims are persecuted by the Government, they still practise bride price (Faye 2021). Although marriage norms are rare today in developed countries, forcibly displaced population may bring their marriage to gain new social networks at the new destination.

In view of the rise in the number of natural disasters and conflicts in many parts of the world, some policy recommendations emerge from this paper. First, I contribute to the current political debate about the impacts of shocks triggering population outflows early in life. Second, by documenting the differential impacts for disaster-induced migrants, this paper helps uncover which policies could potentially offer a more cost-effective way to respond to their migration. For example, a future cash transfer program aimed at decreasing early marriage after a displacement would be more efficient if targeted, not only below 18, but until 22 years old. Third, my findings highlight the importance of culture in shaping displaced populations' economic behaviour. Understanding the role of cultural norms can contribute to effective policy design and evaluation.

Appendix Chapter 1

A Additional Descriptive Figures and Tables

Figure A.1.1: Distribution of the age at marriage, for women



Note: This Figure presents the distribution of ages at first marriage for women in Indonesia from the Indonesia Family Life Survey (IFLS). Non-married women are not included in a category but in the denominator of the calculation of these percentages.

Figure A.1.2: Dataset example

Woman i exposed to an earthquake in 01/2001 and moved right after, got married at age 15

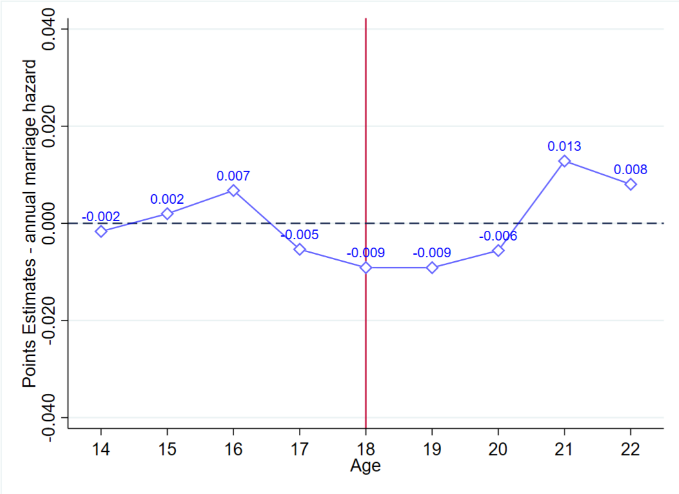
Year	**Month**	**Age**	**Marriage**	**Eq subdistrict**	**Earthquake**	**Migration**
1999	1	12	0	1	0	0
1999	...	12	0	1	0	0
2000	1	13	0	1	0	0
2000	...	13	0	1	0	0
2001	1	14	0	1	1	1
2001	...	14	0	1	1	1
2002	1	15	1	1	1	1

After age 15, woman i exits the dataset

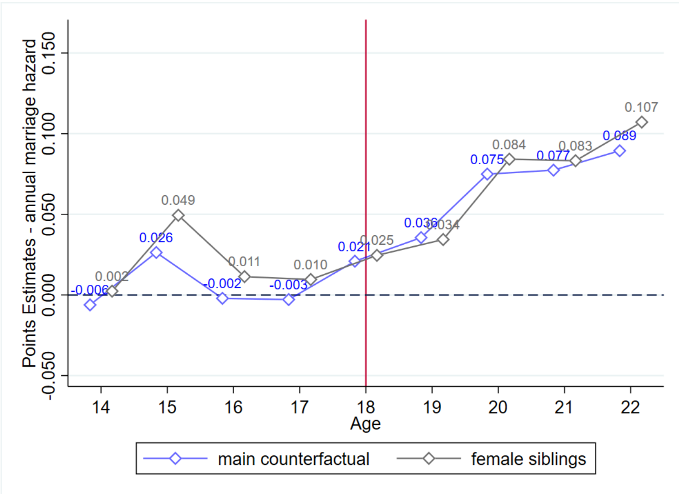
Note: This figure shows a simplified example of the data-set structure. Observations are at the level of the person's age at the month level (from 12 to 22 or the age of first marriage). However, I present an example with a yearly variation to simplify the illustration. The dependent variable, *Marriage*, is a binary variable for marriage, coded to one if the woman married at the age corresponding to the observation. The treatment, *Earthquake*, switches to one if there is an earthquake at their sub-district of residence the year-month corresponding to the observation. If a woman migrates right after an earthquake, *Migration* switches to one.

Figure A.1.3: Effects on timing of marriage, by women age

a) Earthquake effects

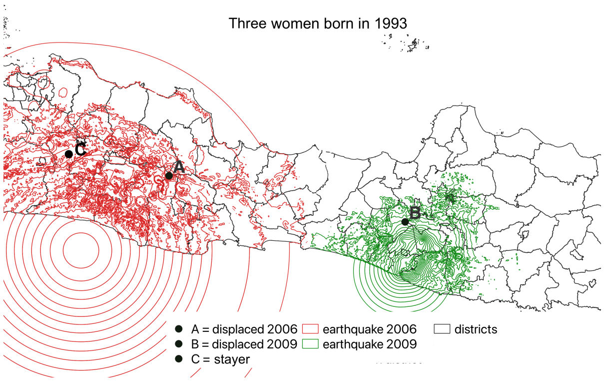


b) Earthquake effects, *earthquake-induced migrants* versus *left-behind women*



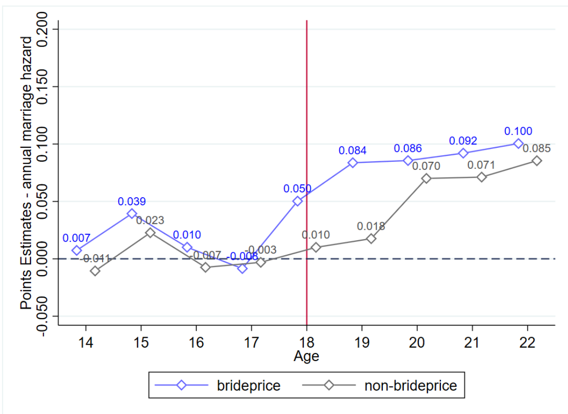
Note: These graphs plot the coefficients obtained from a regression of the annual marriage hazard on the interaction between earthquake's exposure in the residence sub-district and women's age. The regressions control for the province, age, year-of-birth, urban fixed effects, religion, and mother education one year before an earthquake. The Y-axis shows the estimated coefficients and the X-axis shows the ages. Data comes from the Indonesia Family Life Survey from 1993 to 2014 and the United States Geological Survey. Standard errors are clustered at the district level. Treatment is defined at the year-month level. Figure A shows the results from equation 1 with all my sample. Figure b shows the results of equation 4 for women exposed to earthquakes, I further control urban at origin fixed-effects to compare *earthquake-induced migrants* versus *left-behind women*. In figure b, I present the results for my main counterfactual in blue (*left-behind women*) and *left-behind* siblings counterfactual in grey.

Figure A.1.4: Counterfactual example, heterogeneity analysis



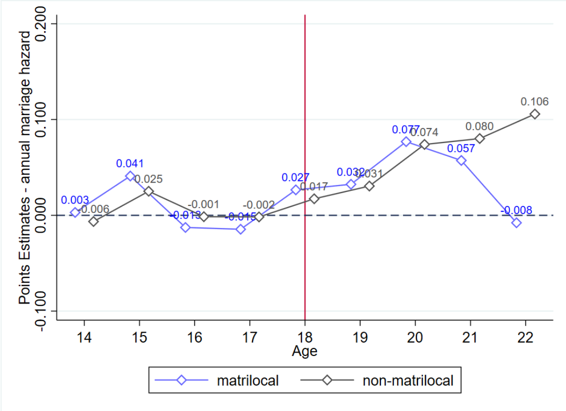
Note: This Figure shows an example of the counterfactual I use to study the heterogenous effects between *earthquake-induced migrants* versus *left-behind women*. I compare women living in the same province, born in the same year, with the same age. Imagine three women from the same province born in 1993. The three of them suffered an earthquake at their place of residence. Women A and C in 2006, and woman B in 2009. Woman A migrated right after the 2006 earthquake, then is called *earthquake-induced migrant women*. Woman C stayed at her place of residence after the 2006 earthquake, then is called *left-behind women*. Women B also migrated right after the 2009 earthquake. Therefore, in 2005 women A, B, and C are non-exposed to earthquakes women. In 2006, woman A was exposed to an earthquake and became an *earthquake-induced migrant woman*. In 2006, woman C was also exposed to an earthquake but a *left-behind women*. Because she stays in the area affected. However, woman B was non-exposed to the earthquake women in 2006.

Figure A.1.5: *Induced migrants* vs *left-behind women*, by age and marriage norms



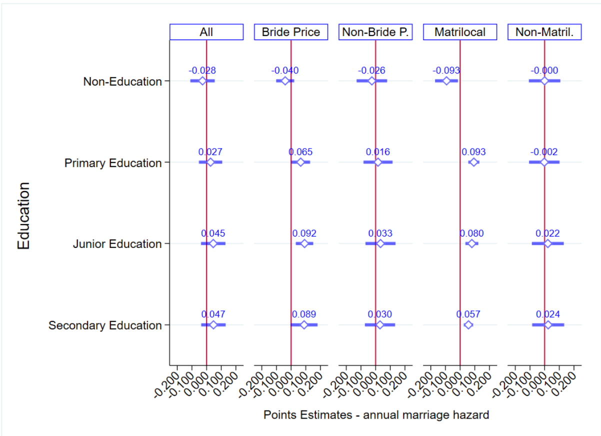
Note: These graphs plot the coefficients obtained from a regression of the annual marriage hazard (equation 4) on the interaction between earthquake exposure in the sub-district of residence and women's age. The Y-axis shows the estimated coefficients and the X-axis shows the ages. I restrict the sample to women exposed to earthquakes to compare *earthquake-induced migrants* versus *left-behind women*. I present the results for bride price women in blue and non-bride price in grey.

Figure A.1.6: *Induced migrants* vs *left-behind women*, by age and kinship tradition



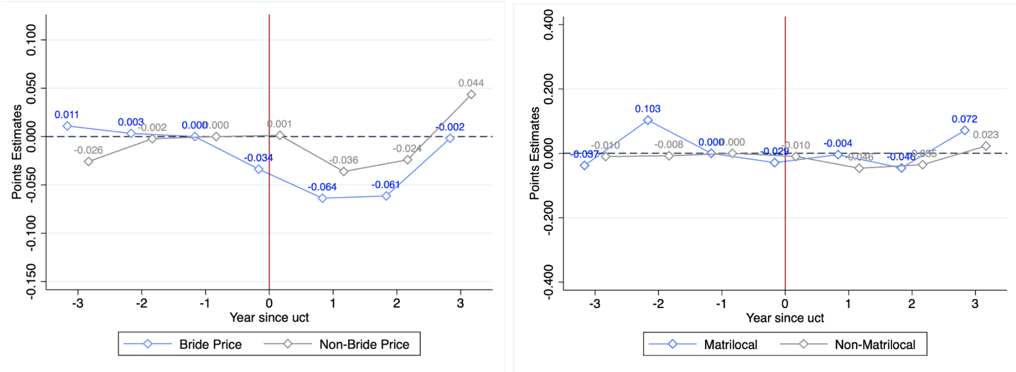
Note: These graphs plot the coefficients obtained from a regression of the annual marriage hazard (equation 4) on the interaction between earthquake exposure in the sub-district of residence and women's age. The Y-axis shows the estimated coefficients and the X-axis shows the ages. I restrict the sample to women exposed to earthquakes to compare *earthquake-induced migrants* versus *left-behind women*. I present the results for matrilocal women in blue and non-matrilocal in grey.

Figure A.1.7: *Induced migrants* vs *left-behind women*, by education and norm



Note: These graphs plot the coefficients obtained from a regression of the annual marriage hazard (equation 4) on the interaction between earthquake exposure in the sub-district of residence and women's education. The X-axis shows the estimated coefficients and the Y-axis shows the education. I restrict the sample to women exposed to earthquakes to compare *earthquake-induced migrants* versus *left-behind women*. From left to right, I present the results for all samples, bride price, non-bride price, matrilocal, and non-matrilocal women. The education is structured in primary (typical ages from 6-7 to 11-12), junior (typical ages from 12-13 to 14-15), and secondary (typical ages from 15-16 to 17-18) education.

Figure A.1.8: Effects of UCT program for *induced migrant women*, by cultural norm



Note: This figure shows the effects of Unconditional Cash Transfer (UCT) on the probability of getting married by year since receiving the transfer. Each figure presents the estimates by cultural norm (bride price and matrilocality). I report coefficient estimates and 95% confidence intervals from a regression of getting married that year on the interaction between an indicator variable for being a uct-beneficiary household from 2005 to 2014 and a variable for years since a household received the transfer for the first time, a time-varying measure of covariates, and province, age, cohort, urban fixed effects (equation (5)). The omitted category is T-1, one year before the transfer. Standard errors are clustered at the district level. The data-set is in a person-year panel format. Treatment is defined at the year level.

Table A.1.1: Destructive earthquakes in Indonesia (1993-2014)

Year	Month	epicenter province	IFLS province affected	MM intensity mean	MM intensity maximum	N exposed N exposed	Sample exposed
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1994	2	Lampung	Banten, Jakarta, West Java Lampung, South Sumatra, North Sumatra	3.8	6.6	7,306	616
1995	10	Jambi	West Sumatra	5	6.8	892	98
2000	6	Bengkulu	Banten, Jakarta, West Java Lampung, West Sumatra, South Sumatra	5	8	5,425	577
2002	11	Aceh	North Sumatra	5	6.8	1,163	43
2004	12	Aceh	West Sumatra, North Sumatra	5	8	4,357	221
2005	3	North Sumatra	Lampung, West Sumatra North Sumatra, South Sumatra	5	8.4	1,204	74
2005	5	Aceh	North Sumatra	5	6.6	52	3
2006	5	Yogyakarta	Yogyakarta, Jakarta, West Java Central Java, East Java	5	7.6	10,501	533
2007	3	West Sumatra	West Sumatra, North Sumatra	5	7	60	5
2007	9	West Sumatra	Lampung, West Sumatra, South Sumatra	5	7	1,752	84
2007	11	West Nusa Tenggara	West Nusa Tenggara	4	6.6	2,463	155
2008	2	Aceh	North Sumatra	5	7	843	20
2008	2	West Sumatra	West Sumatra	5	6.8	783	36
2008	5	North Sumatra	West Sumatra, North Sumatra	5	6.6	1,658	68
2009	9	West Java	Banten, Jakarta, West Java Central Java, Lampung	5	6.8	16,922	570
2009	9	West Sumatra	West Sumatra, South Sumatra North Sumatra	5	8.2	3,494	138
2009	10	Bengkulu	Lampung, West Java, South Sumatra	4	6.8	408	13
2010	4	Aceh	West Sumatra, North Sumatra	5	7	1,298	27
2010	10	West Sumatra	West Sumatra, South Sumatra	5	7	25	1
2011	2	South Sulawesi	South Sulawesi	5	7.2	1,675	77
2013	7	Aceh	North Sumatra	5	6.6	20	0
Total	21	10 prov. epic.	21 prov. affected			62,306	3,359

Note: This Table presents the descriptive statistics of the large earthquakes that this paper exploits for identification. Large earthquakes are defined as earthquakes with an intensity of at least VII in some of its locations affected(Gignoux and Menéndez 2016). There are 21 earthquakes in total. Columns (2) and (3) report the year and month of each earthquake. Column (3) presents the province where the epicentre falls. Column (4) includes the provinces within the IFLS survey affected by each earthquake. Columns (5) and (6) report the mean and maximum value of an earthquake ground shaking (measured in Modified Mercalli intensity unit). Column (7) presents the population living in the sub-district affected when the earthquake occurs, and, Column (8) the women in my sample.

Table A.1.2: Sample: Descriptive statistics

variable	**N**	**mean**	**sd**
Individual characteristics			
age at marriage	17,731.0	20.8	3.7
primary education	29,017.0	0.7	0.5
secondary education	29,017.0	0.2	0.4
employed	22,906.0	0.3	0.5
muslim	29,278.0	0.9	0.3
javanese	30,166.0	0.4	0.5
bride price	29,916.0	0.3	0.4
matrilocality	6,404.0	0.8	0.4
Father characteristics			
age	16,540.0	46.9	11.2
at least primary educ.	17,243.0	0.4	0.5
employed	1,288.0	0.1	0.3
self-employed	199.0	0.2	0.4
agriculture sector	5,820.0	0.4	0.5
Mother characteristics			
age	18,399.0	41.8	10.3
at least primary	17,898.0	0.5	0.5
married	17,996.0	0.9	0.3
employed	7,400.0	0.1	0.3
number siblings	30,892.0	0.7	1.5
female siblings	30,892.0	0.5	1.0
Household characteristics			
participation *arisan*	30,892.0	0.5	0.5
participation associations	30,892.0	1.3	2.0
house property	28,630.0	0.8	0.4
farm property	28,630.0	0.2	0.4
livestock property	28,630.0	0.1	0.4
labour income	15,945.0	4,690,546	2.4
non-labour income	28,625.0	770,197.4	1.5
cement wall	29,530.0	0.7	0.5
concrete roof	30,892.0	0.0	0.1

Note: This Table presents the descriptive statistics of my sample. I restrict my sample to women that are at least 23 years old at the last interview.

Table A.1.3: Differences in characteristics between exposed and non-exposed women

	Mean	**Mean**	**Diff (2) - (1)**
	Non-Exposed	Exposed	
Individual Characteristics			
age	-0.814	-0.682	0.130***
	(0.316)	(0.326)	(0.010)
non-muslim	0.141	-0.119	-0.261***
	(1.135)	(0.851)	(0.028)
non-javanese	0.024	0.016	-0.008
	(0.994)	(0.996)	(0.030)
married	-0.688	-0.523	0.164***
	(0.587)	(0.767)	(0.022)
at least primary	-0.380	0.025	0.396***
	(0.943)	(0.999)	(0.029)
employed	-0.664	-0.537	0.125***
	(0.633)	(0.764)	(0.025)
siblings	0.461	0.034	-0.426***
	(1.183)	(1.118)	(0.034)
female siblings	0.932	0.438	-0.493***
	(1.413)	(1.287)	(0.040)
Parent Characteristics			
age father	-0.200	-0.064	0.135***
	(0.812)	(0.788)	(0.028)
at least primary father	-0.144	-0.090	0.046
	(0.983)	(0.992)	(0.033)
working father	0.125	-0.042	-0.187
	(1.163)	(0.941)	(0.143)
self-employed father	-0.194	-0.055	0.134
	(0.930)	(0.998)	(0.366)
in agriculture father	0.093	0.085	0.050
	(1.027)	(1.023)	(0.062)
age mother	-0.312	-0.161	0.150***
	(0.691)	(0.701)	(0.024)
at least primary mother	-0.162	-0.076	0.075**
	(1.011)	(1.008)	(0.033)
Household Characteristics			
arisan participation	-0.190	0.237	0.428***
	(0.923)	(1.041)	(0.030)
com. organization part.	-0.218	0.140	0.357***
	(0.843)	(1.154)	(0.032)
own house	0.127	-0.023	-0.133***
	(0.895)	(1.016)	(0.029)
own farm	-0.036	0.007	0.048
	(0.970)	(1.006)	(0.030)
own livestock	0.286	0.009	-0.263***
	(1.233)	(1.010)	(0.032)
value own house	-0.107	-0.018	0.080***
	(0.840)	(0.907)	(0.027)
value farm	-0.047	-0.039	0.007
	(0.590)	(0.758)	(0.022)
savings	0.059	0.002	-0.061
	(1.751)	(0.972)	(0.038)
labour income	0.079	-0.029	-0.115**
	(2.120)	(0.355)	(0.047)
non-labour income	0.010	-0.026	-0.038
	(1.391)	(0.437)	(0.025)
Observations	1,584	3,359	4,943

Note: This table shows along which dimensions *exposed)* and *non-exposed* women differ. I report coefficient estimates together with 95% confidence intervals from a regression of an indicator variable for earthquake exposure at baseline on socio-economic characteristics before an earthquake and urban fixed effects. *** p<0.01, ** p<0.05, * p<0.1

Table A.1.4: Effects of earthquakes on migration decisions

	migration				marriage migration	migration
	(1)	(2)	(3)	(4)	(5)	(6)
Earthquake	0.138***	0.139***	0.135***	0.135***	-0.022***	0.265***
	(0.008)	(0.009)	(0.011)	(0.009)	(0.002)	(0.021)
*Earthquake * Women*		-0.006*				
		(0.003)				
*Earthquake * Non-Javanes*			0.002			
			(0.012)			
*Earthquake * Non-Muslim*				0.005		
				(0.014)		
*Earthquake * years to 23*						-0.006***
						(0.001)
Dep. var. mean (1993)	0.120	0.120	0.120	0.120	0.120	0.120
Observations	94,329	94,329	94,329	94,329	237,726	16,284
Number of provinces	15	15	15	15	15	15
Number of survey years	5	5	5	5	5	5
Number of districts	255	255	255	255	255	255
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Year Survey FE	Yes	Yes	Yes	Yes	Yes	Yes
Urban FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Note: This Table presents the estimates from Equation 3 with a refined migration definition. The dependent variable is a binary variable for migration right after an earthquake, coded to one of the individuals migrate from their place of residence. Namely, it is a migration that happens right after the occurrence of an earthquake. In this table, I include the migration that takes place during the 24 months after an earthquake. In Appendix Table A.20, I conduct the same analysis restricting my migration window from 14 to 6 months. Results hold too. Earthquakes are defined as earthquakes with an intensity of at least VII in some of its locations affected (Gignoux and Menéndez 2016). Observations are at the survey year level. Column (1) presents the main results. Columns (2), (3) and (4) report the heterogeneity results by gender, ethnicity, and religion. In Indonesia, 43% is Javanese and 87% Muslim (Population Census, 2010). Column (5) presents the results for migration as a consequence of marriages (*marriage migration*). Column (6) shows the results for women below 23 in an interaction with the age gap of 23. Standard errors are clustered at the district level. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.1.5: Destination of *earthquake-induced migrant women*, by origin

IFLS PROVINCES	within same desa	within same kec.	within same kab.	within same prov.	other IFLS prov.	other prov.
	(1)	(2)	(3)	(4)	(5)	(6)
Sumatra	29.70%	20.79%	13.86%	16.83%	13.53%	5.28%
North Sumatra	21.36%	25.24%	20.39%	19.42%	9.71%	3.88%
South Sumatra	32.10%	23.46%	16.05%	8.64%	12.35%	7.41%
West Sumatra	36.62%	12.68%	5.63%	25.35%	18.31%	1.41%
Lampung	33.33%	18.75%	8.33%	12.50%	16.67%	10.42%
Java	48.23%	7.08%	13.27%	15.04%	14.16%	2.21%
Banten	37.84%	10.81%	13.51%	10.81%	24.32%	2.70%
D.I Yogyakarta	60.00%	5.00%	0.0%	20.00%	15.00%	0.0%
DKI Jakarta	86.67%	6.67%	0.0%	0.0%	6.67%	0.0%
West Java	52.83%	5.66%	13.21%	16.98%	7.55%	3.77%
Central Java	35.38%	10.77%	16.92%	13.85%	21.54%	1.54%
East Java	52.78%	0.0%	19.44%	22.22%	2.78%	2.78%
Nussa Teggara	30.30%	36.36%	15.15%	9.09%	9.09%	0.0%
Indonesia	38.29%	16.80%	13.57%	15.35%	12.60%	3.39%

Note: This Table presents the destination of *earthquake-induced migrant* women by provinces of origin with the Indonesia Family Life Survey (IFLS).

Table A.1.6: Transfer at the moment of the marriage, *induced migrant* vs *left-behind*

VARIABLES	bride price value	bp value	bp value	bp value	bp value	bp value
	(1)	(2)	(3)	(4)	(5)	(6)
	Full sample			Bride Price subsample		
	PANEL A: Contemporary price- All					
Earthquake	0.098**	-0.362**	-0.098	-0.057	-0.745***	-0.934***
	(0.044)	(0.172)	(0.236)	(0.083)	(0.241)	(0.346)
*Earthquake * Migration*	0.004	0.024	0.024	0.302*	0.399***	0.551***
	(0.074)	(0.074)	(0.081)	(0.168)	(0.146)	(0.190)
Observations	295,044	282,636	219,852	82,980	79,104	61,968
	PANEL B: Contemporary price- Bellow 18					
Earthquake	0.124***	0.153	0.280	0.019	-0.107	-0.147
	(0.041)	(0.182)	(0.281)	(0.068)	(0.145)	(0.236)
*Earthquake * Migration*	-0.113*	-0.123*	-0.103	0.040	0.083	0.107
	(0.068)	(0.069)	(0.074)	(0.189)	(0.189)	(0.269)
Observations	190,824	180,408	142,320	54,276	50,940	40,584
Province/ Urban FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth Yr/ Age FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	No	Yes	Yes
	PANEL C: Descomposition by groom					
	all	migrant	native	all	migrant	native
Earthquake	-0.051	1.079	0.030	-0.970	-1.753*	-0.679
	(0.531)	(0.936)	(0.541)	(0.769)	(0.843)	(0.860)
*Earthquake * Migration*	0.019	0.256	0.008	0.107	2.014*	-1.106**
	(0.149)	(0.310)	(0.174)	(0.337)	(1.112)	(0.473)
Observations	36,816	5,520	31,296	11,304	1,368	9,936

Note: This Table presents the estimates from Equation 4 where *Earthquake * Migration* is the interaction of earthquakes with a migration right after an earthquake. The dependent variable is a continuous variable for payment at marriage (*bride payment*) at the age of marriage. Earthquakes are defined as earthquakes with an intensity of at least VII in some of its locations affected (Gignoux and Menéndez 2016). The counterfactual for the *Earthquake * Migration* interaction is the *left-behind women*. Observations are at the woman level. Panel A reports the results on the timing of marriage. Columns (1) to (3) show the results for the full sample. Column (1) presents the results without covariates. Column (2) includes covariates (religion and mother's education before an earthquake) and woman's education fixed effects. Column (3) also controls for the spouse's age at marriage. Columns (4) to (6) show the results for a sub-sample of women traditionally engaged in the bride price custom. Panel B presents the same analysis but for marriages below 17. Panel C shows the results by sub-sample of grooms': *migrant* or *native*. *Migrant* is a man that suffers an earthquake and migrates right after an earthquake. I define *native* as a groom that is not classified as *migrant*. The number of observations decreases because not every woman has data on her spouse. Standard errors are clustered at the district level. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.1.7: Groom's education at marriage, *induced migrant* vs *left-behind*

VARIABLES	educ. spouse	educ. spouse	educ. spouse	educ. spouse	educ. spouse	educ. spouse
	(1)	(2)	(3)	(4)	(5)	(6)
	Below age 23			Below age 18		
	all	matrilocal	non-matrilocal	all	matrilocal	non-matrilocal
Earthquake	0.601***	2.085***	0.654***	0.606***	-0.813***	0.695***
	(0.159)	(0.631)	(0.158)	(0.179)	(0.223)	(0.184)
*Earthquake * Migration*	0.090	0.750***	0.011	-0.031	0.578**	-0.100
	(0.108)	(0.202)	(0.117)	(0.123)	(0.242)	(0.134)
Observations	62,640	5,304	56,928	50,064	4,260	45,420
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Urban FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Note: This Table shows the estimates from Equation 4 where *Earthquake * Migration* is the interaction of earthquakes with a migration right after an earthquake. The dependent variable is a continuous variable for the education gap between spouses at marriage. Earthquakes are defined as earthquakes with an intensity of at least VII in some of its locations affected (Gignoux and Menéndez 2016). Therefore, the counterfactual is *left-behind* women. Observations are at the woman level. The estimation includes province, age, year of birth, urban at the origin of residence, woman's education fixed effects, and covariate (religion). Column (1) includes the entire sample. Column (2) restricts the sample to women traditionally engaged in a matrilocal custom, and Column (3) engaged in non-matrilocal customs. Columns (4), (5) and (6) perform the same analysis that Columns (1), (2) and (3) but for marriages below 17. Standard errors are clustered at the district level. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.1.8: Unconditional Cash Transfer: effect on the timing of marriage

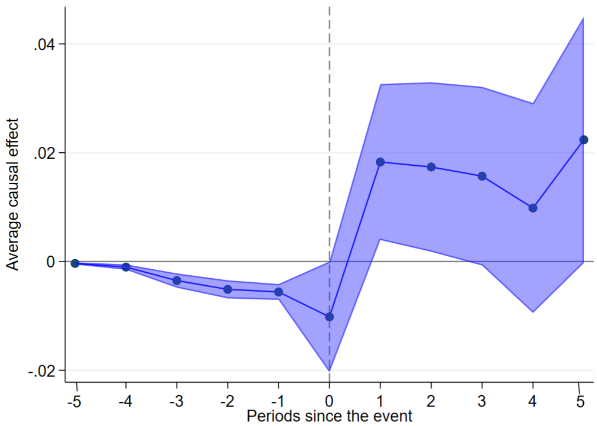
	(1)	(2)	(3)	(4)	(5)	(6)
	PANEL A: Unmatching			PANEL B: Coarsened Exact Matching		
	UCT interaction	UCT-hhs	NonUCT-hhs	UCT interaction	UCT-hhs	NonUCT-hhs
*Earthquake * Migration*	0.029***	0.009	0.032**	0.036***	-0.000	0.040***
	(0.002)	(0.008)	(0.008)	(0.013)	(0.018)	(0.013)
*Earthquake * Migration * UCT*	-0.008			-0.020		
	(0.014)			(0.017)		
Observations	107,316	22,368	84,948	101,664	21,564	80,100
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Urban FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table displays the estimation results for the effect of earthquakes on the timing of marriage with two interactions: Migration interaction and Unconditional Cash Transfers (UCT) interaction. In Panel A, I present the unmatching estimates, and in Panel B, the Coarsened Exact Matching estimates. In Columns (1) and (4), I regress an indicator variable that takes value 1 when a woman gets married for the first time (0, otherwise) on *Earthquake*, *Earthquake * Migration*, and *Earthquake * Migration* interacted to a variable equal to 1 if a woman's household is beneficiary of an UCT; a time-varying measure of covariates, province fixed effects, age fixed effects, year-of-birth fixed effects, and urban fixed-effects (equation (5)). In Columns (2), (3), (5) and (6), I estiamte (equation (4)) by a sumsample of UCT-beneficiaries and nonbeneficiaries. Standard errors are clustered at the district level. The dataset is in a person-age panel format. Treatment is defined at the year level. Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

B Robustness checks

B.1 Earthquake's effects

Figure A.1.9: Effect of earthquakes on the timing of marriage, by year since treatment



Note: This figure plots the event and year coefficient from estimating equation 2 using the timing of marriage as the dependent variable. The confidence intervals are 95%. Marriage outcomes come from the IFLS and earthquake variation from USGS. The omitted category is T-1, earthquake year. The data-set is a person-age panel format. Treatment is defined at the year level. Figure A.8 present the estimates for Chaisemartin and D'Haultfœuille 2020 estimator. Similar estimates for the Callaway and PedroSant'Anna 2021 and Sun and Abraham 2021 estimators.

Table A.1.9: Earthquakes effects on the timing of marriage, lag effects

	(1)	(2)	(3)	(4)	(5)	(6)
$Earthquake_{year_0}$	-0.003					
	(0.006)					
$Earthquake_{year_1}$		0.005				
		(0.006)				
$Earthquake_{year_2}$			0.007*			
			(0.004)			
$Earthquake_{year_3}$				0.007*		
				(0.004)		
$Earthquake_{year_4}$					0.005	
					(0.003)	
$Earthquake_{year_5}$						0.004
						(0.003)
Observations	585,816	585,816	585,816	585,816	585,816	585,816
$Earthquake_{year_6}$	0.006**					
	(0.003)					
$Earthquake_{year_7}$		0.006**				
		(0.003)				
$Earthquake_{year_8}$			0.005*			
			(0.003)			
$Earthquake_{year_9}$				0.005*		
				(0.003)		
$Earthquake_{year_{10}}$					0.006**	
					(0.003)	
$Earthquake_{year_{11}}$						0.006*
						(0.003)
Observations	585,816	585,816	585,816	585,816	585,816	585,816
Dep. var. mean	0.036	0.036	0.036	0.018	0.018	0.018
Number of provinces	15	15	15	15	15	15
Number of years	22	22	22	22	22	22
Number of districts	255	255	255	255	255	255

Note: This Table presents the earthquake results on the dependent variable: annual marriage hazard. The dependent variable is a binary variable for marriage, coded to one if the woman married at the age corresponding to the observation. I provide different earthquakes treatments based on the lag effects from 0 to 11 years after an earthquake. Observations are at the level of person age at month level (from 12 to 22 or age of first marriage). The baseline specification is presented in Equation 1. Columns (1) to (6) present the results with province, age, birth year, and urban fixed effects, and controls for baseline characteristics (religion and mother education for the year before earthquake). Standard errors are clustered at district level. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.1.10: Earthquakes effects on the timing of marriage, alternative definitions

	Below age 23			Below age 18		
	(1)	(2)	(3)	(4)	(5)	(6)
	PANEL A: Earthquakes before 12					
Earthquake	0.016***	0.008**	0.007**	0.007**	-0.000	-0.000
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
Observations	585,816	585,816	585,816	350,232	350,232	350,232
	PANEL B: Multiple earthquakes					
Earthquake	0.010***	0.008**	0.007**	0.010***	0.009***	0.008**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Observations	585,816	585,816	585,816	350,232	350,232	350,232
Dep. var. mean	0.036	0.036	0.036	0.018	0.018	0.018
Number of provinces	15	15	15	15	15	15
Number of years	22	22	22	22	22	22
Number of districts	255	255	255	255	255	255
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Urban FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth Year FE	No	Yes	Yes	No	Yes	Yes
Age FE	No	Yes	Yes	No	Yes	Yes
Controls	No	No	Yes	No	No	Yes

Note: This Table presents the earthquake results on the dependent variable: annual marriage hazard. The dependent variable is a binary variable for marriage, coded to one if the woman married at the age corresponding to the observation. Panel A includes all destructive earthquakes in a woman life (from age 0 to age 22). Panel B reports the analysis for a continuous definition of treatment, to capture multiple earthquakes. Observations are at the level of person age at month level (from 12 to 22 or age of first marriage). The baseline specification is presented in Equation 1. Column (1) presents the results without age, birth year fixed effects an covariates. Column (2) includes age and birth year fixed effects. Column (3) controls for baseline characteristics (religion and mother education for the year before earthquake). Columns (4), (5) and (6) perform the same analysis that Columns (1), (2) and (3) but for a sub-sample of ages from 12 to 17 (or age of first marriage). Standard errors are clustered at district level. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.1.11: Earthquakes effects on spouse cohabitation

	Below age 23			Below age 18		
VARIABLES	cohabitation	cohabitation	cohabitation	cohabitation	cohabitation	cohabitation
	(1)	(2)	(3)	(4)	(5)	(6)
Earthquake	0.002**	0.002*	0.002**	0.002***	0.002***	0.002***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	525,156	585,816	525,156	350,232	350,232	350,232
Dep. var. mean	0.036	0.036	0.036	0.018	0.018	0.018
Number of provinces	15	15	15	15	15	15
Number of years	22	22	22	22	22	22
Number of districts	255	255	255	255	255	255
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Urban FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth Year FE	No	Yes	Yes	No	Yes	Yes
Age FE	No	Yes	Yes	No	Yes	Yes
Controls	No	No	Yes	No	No	Yes

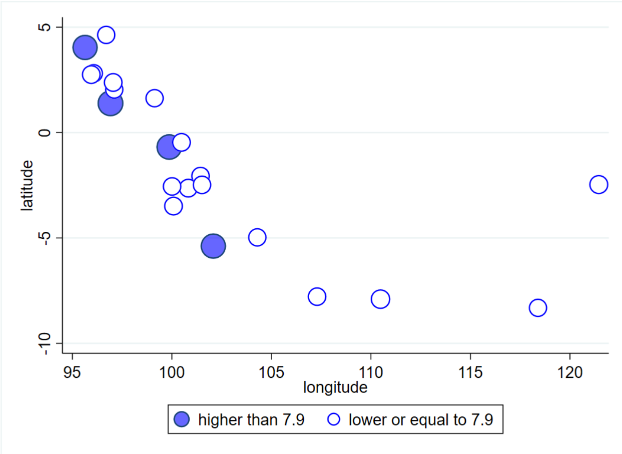
Note: This Table presents the earthquake results on the dependent variable: spouse cohabitation. The dependent variable is a binary variable for a change of household after marriage, coded to one if the woman move to another household after marriage. In this analysis I remove matrilocal women. Observations are at the level of person age at month level (from 12 to 22 or age of first marriage). The baseline specification is presented in Equation 1. Column (1) presents the results without age, birth year fixed effects an covariates. Column (2) includes age and birth year fixed effects. Column (3) controls for baseline characteristics (religion and mother education for the year before earthquake). Columns (4), (5) and (6) perform the same analysis that Columns (1), (2) and (3) but for a sub-sample of ages from 12 to 17 (or age of first marriage). Standard errors are clustered at district level. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.1.12: Earthquakes effects on timing of marraige, excluding arranged marriages

	Below age 23			Below age 18		
VARIABLES	getting married	get. married	get. married	get. married	get. married	get. married
	(1)	(2)	(3)	(4)	(5)	(6)
Earthquake	0.010***	0.008**	0.007**	0.010***	0.009***	0.008**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Observations	585,600	585,600	585,600	350,148	350,148	350,148
Dep. var. mean	0.036	0.036	0.036	0.018	0.018	0.018
Number of provinces	15	15	15	15	15	15
Number of years	22	22	22	22	22	22
Number of districts	255	255	255	255	255	255
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Urban FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth Year FE	No	Yes	Yes	No	Yes	Yes
Age FE	No	Yes	Yes	No	Yes	Yes
Controls	No	No	Yes	No	No	Yes

Note: This Table presents the earthquake results on the dependent variable: annual marriage hazard. The dependent variable is a binary variable for marriage, coded to one if the woman married at the age corresponding to the observation. For this analysis, I restrict my sample to marriages that are not arranged marriages. Observations are at the level of a person's age at the month level (from 12 to 22 or the age of first marriage). The baseline specification is presented in Equation 1. Column (1) presents the results without age, birth year fixed effects and covariates. Column (2) includes age and birth year fixed effects. Column (3) controls for baseline characteristics (religion and mother education for the year before the earthquake). Columns (4), (5) and (6) perform the same analysis that Columns (1), (2) and (3) but for a sub-sample of ages from 12 to 17 (or age of first marriage). Standard errors are clustered at the district level. Robust standard errors are in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Figure A.1.10: Earthquake-epicenters coordinates in Indonesia from 1994 to 2014



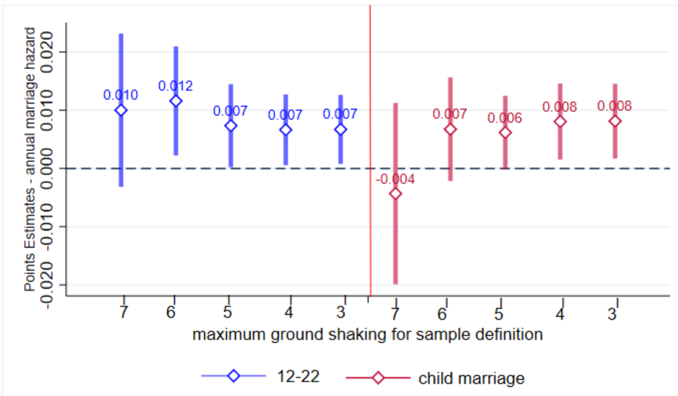
Note: This Figure shows the longitude and the latitude of the 21 earthquakes I exploit for identification. In darker blue, the earthquakes with a magnitude of 8 or higher.

Table A.1.13: Earthquakes effects in areas highly exposed to earthquakes

	Below age 23			Below age 18		
VARIABLES	getting married	get. married	get. married	get. married	get. married	get. married
	(1)	(2)	(3)	(4)	(5)	(6)
PANEL A: Baseline specification						
Earthquake	0.010***	0.008**	0.007**	0.010***	0.009***	0.008**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Observations	585,816	585,816	585,816	350,232	350,232	350,232
PANEL B: The westest zone - Sumatra island sample						
Earthquake	0.008***	0.005*	0.005	0.009***	0.007**	0.007**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)
Observations	534,096	534,096	534,096	318,084	318,084	318,084
PANEL C: The closest provinces to seismic plaques						
Earthquake	0.013***	0.011**	0.010**	0.009**	0.007*	0.006*
	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.003)
Observations	144,912	144,912	144,912	86,040	86,040	86,040
PANEL D: The closest districts to seismic plaques						
Earthquake	0.013**	0.012*	0.012*	0.017*	0.015	0.016*
	(0.007)	(0.007)	(0.007)	(0.008)	(0.009)	(0.008)
Observations	126,948	126,948	126,948	74,868	74,868	74,868
Dep. var. mean	0.036	0.036	0.036	0.018	0.018	0.018
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Urban FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth Year FE	No	Yes	Yes	No	Yes	Yes
Age FE	No	Yes	Yes	No	Yes	Yes
Controls	No	No	Yes	No	No	Yes

Note: This Table shows alternative samples for the difference-in-difference regressions reported in table 1. This table provides three different definitions of areas highly exposed to earthquakes: Sumatra island, the western island (in Panel A); provinces closer to seismic plaques; and districts (in Panel B); closest to seismic plaques (in Panel C). Observations are at the level of person × age (from 12 to 22 or age of first marriage, whichever is earlier). *** p<0.01, ** p<0.05, * p<0.1

Figure A.1.11: Alternative earthquake definitions, effects on the timing of marriage



Note: This Figure shows the effects of destructive earthquakes on the timing of marriage for a set of alternative definitions of destructive earthquakes. I use the maximum ground-shaking earthquakes generated on an island to provide five different definitions. Estimates remain significant and positive. In red are the child marriage hazard estimates. The estimates are from equation (1).

Table A.1.14: P-values for alternative clustering methods for table 1

	cluster at district level	cluster at province level	cluster at island level	bootstrap cluster
	(1)	(2)	(3)	(4)
		Bellow age 23		
Column 1	(0.003)	(0.003)	(0.001)	(0.001)
Column 2	(0.003)	(0.003)	(0.002)	(0.001)
Column 3	(0.003)	(0.003)	(0.001)	(0.001)
		Bellow age 18		
Column 4	(0.003)	(0.002)	(0.001)	(0.001)
Column 5	(0.003)	(0.003)	(0.001)	(0.001)
Column 6	(0.003)	(0.003)	(0.002)	(0.001)

Note: This table shows p-values for the difference-in-difference regressions reported in table 1 for the full regression samples: women aged 23 or older at the last interview. Observations are at the level of person $\times$ age (from 12 to 22 or age of first marriage, whichever is earlier). All regression specifications include island x year, urban, age and year of birth fixed effects. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A.1.15: Alternative baseline specification for Table 1

	Below age 23			Below age 18		
VARIABLES	getting married	get. married	get. married	get. married	get. married	get. married
	(1)	(2)	(3)	(4)	(5)	(6)
PANEL A: Baseline specification						
Earthquake	0.010***	0.008**	0.007**	0.010***	0.009***	0.008**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Observations	585,816	585,816	585,816	350,232	350,232	350,232
PANEL B: Island fixed effects instead of Province						
Earthquake	0.020***	0.008***	0.007**	0.012***	0.008**	0.008**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Observations	585,816	585,816	585,816	350,232	350,232	350,232
PANEL C: District fixed effects instead of Province						
Earthquake	0.026***	0.010***	0.010***	0.013***	0.008***	0.012***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)
Observations	585,816	585,816	585,816	350,232	350,232	350,232
PANEL D: Sub-district clustered instead of at district level						
Earthquake	0.010***	0.008***	0.007**	0.010***	0.009***	0.008***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Observations	582,624	582,624	582,624	347,892	347,892	347,892
Dep. var. mean	0.036	0.036	0.036	0.018	0.018	0.018
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Urban FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth Year FE	No	Yes	Yes	No	Yes	Yes
Age FE	No	Yes	Yes	No	Yes	Yes
Controls	No	No	Yes	No	No	Yes

Note: This Table shows alternative specifications for the difference-in-difference regressions reported in table 1. Observations are at the level of person × age (from 12 to 22 or age of first marriage, whichever is earlier). *** p<0.01, ** p<0.05, * p<0.1

Table A.1.16: Alternative covariates for Table 1

	Below age 23			Below age 18		
VARIABLES	getting married	get. married	get. married	get. married	get. married	get. married
	(1)	(2)	(3)	(4)	(5)	(6)
Earthquake	0.007**	0.007*	0.008*	0.009**	0.002	0.003
	(0.003)	(0.004)	(0.004)	(0.003)	(0.006)	(0.006)
Observations	555,684	495,984	469,416	332,448	295,788	280,200
Dep. var. mean	0.036	0.036	0.036	0.018	0.018	0.018

Note: This Table includes alternative covariates (father education and number of siblings before an earthquake) to the difference-in-difference regressions reported in Columns (3) and (6) of Table 1. Observations are at the level of person × age (from 12 to 22 or age of first marriage, whichever is earlier). *** p<0.01, ** p<0.05, * p<0.1

Table A.1.17: Data-set at survey level, earthquake's effects on the age at marriage

	Below age 23			Below age 18		
VARIABLES	age at marriage	age at marriage	age at marriage	age at marriage	age at marriage	age at marriage
	(1)	(2)	(3)	(4)	(5)	(6)
Earthquake	-0.399	-0.343**	-0.311**	-0.568*	-0.608**	-0.561**
	(0.282)	(0.156)	(0.145)	(0.328)	(0.253)	(0.233)
Observations	24,540	24,528	24,528	15,818	15,818	15,818
Dep. var. mean	0.036	0.036	0.036	0.018	0.018	0.018
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Urban FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth Year FE	No	Yes	Yes	No	Yes	Yes
Age FE	No	Yes	Yes	No	Yes	Yes
Controls	No	No	Yes	No	No	Yes

Note: This Table shows the difference-in-difference regressions reported in table 1 for a data-set structured at survey level. Observations are at the level of person × age (from 12 to 22 or age of first marriage, whichever is earlier). *** p<0.01, ** p<0.05, * p<0.1

Table A.1.18: Effects of earthquakes on marriage timing, without voluntary migrants

	Below age 23			Below age 18		
VARIABLES	getting married	get. married	get. married	get. married	get. married	get. married
	(1)	(2)	(3)	(4)	(5)	(6)
Earthquake	0.007**	0.005	0.004	0.010***	0.009**	0.008**
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Observations	383,292	383,292	383,292	222,804	222,804	222,804
Dep. var. mean	0.036	0.036	0.036	0.018	0.018	0.018
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Urban FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth Year FE	No	Yes	Yes	No	Yes	Yes
Age FE	No	Yes	Yes	No	Yes	Yes
Controls	No	No	Yes	No	No	Yes

Note: This Table shows alternative specifications for the difference-in-difference regressions reported in table 1. Observations are at the level of person × age (from 12 to 22 or age of first marriage, whichever is earlier). *** p<0.01, ** p<0.05, * p<0.1

Table A.1.19: Effects of earthquakes on the timing of marriage, without returnees

	Below age 23			Below age 18		
VARIABLES	getting married	get. married	get. married	get. married	get. married	get. married
	(1)	(2)	(3)	(4)	(5)	(6)
Earthquake	0.010***	0.007**	0.006*	0.010***	0.009**	0.008**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Observations	578,040	578,040	578,040	348,012	348,012	348,012
Dep. var. mean	0.036	0.036	0.036	0.018	0.018	0.018
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Urban FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth Year FE	No	Yes	Yes	No	Yes	Yes
Age FE	No	Yes	Yes	No	Yes	Yes
Controls	No	No	Yes	No	No	Yes

Note: This Table shows alternative specifications for the difference-in-difference regressions reported in table 1. Observations are at the level of person × age (from 12 to 22 or age of first marriage, whichever is earlier). *** p<0.01, ** p<0.05, * p<0.1

C *Earthquake-induced migrants* versus *left-behind women*

Table A.1.20: *Earthquake-induced migrant* vs *left-behind*, by migration window

VARIABLES	getting married	getting married	getting married	getting married	getting married	getting married
Migration Window	24 months	14 months	12 months	10 months	8 months	6 months
	(1)	(2)	(3)	(4)	(5)	(6)
Earthquake	-0.020***	-0.015**	-0.017***	-0.017**	-0.016**	-0.017**
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
*Earthquake * Migration*	0.027***	0.031***	0.033***	0.029***	0.022***	0.021**
	(0.005)	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)
Observations	227,088	177,708	175,212	170,796	169,464	167,904
Dep. var. mean	0.036	0.036	0.036	0.018	0.018	0.018
Number of provinces	15	15	15	15	15	15
Number of years	22	22	22	22	22	22
Number of districts	255	255	255	255	255	255

Note: This Table presents the estimates from Equation 4 where $Eq_{s,t} * Disp_{s,t}$ is the interaction of earthquakes with a migration right after an earthquake. The dependent variable is a binary variable for marriage, coded to one if the woman married at the age corresponding to the observation. I show the results by migration window, from 24 months to 6 months after an earthquake. Earthquakes are defined as earthquakes with an intensity of at least VII in some of its locations affected (Gignoux and Menéndez 2016). Observations are at the level of a person's age at the month level (from 12 to 22 or the age of first marriage). Columns (1) to (6) present the results with the province, urban at origin, age and birth year fixed effects, and controls for baseline characteristics (religion and mother education for the year before the earthquake). Standard errors are clustered at the district level. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.1.21: Effects of earthquakes on migration decisions, by migration window

VARIABLES	migration	migration	migration	migration	migration	migration
Migration Window	24 months	14 months	12 months	10 months	8 months	6 months
	(1)	(2)	(3)	(4)	(5)	(6)
Earthquake	0.138***	0.043***	0.040***	0.028***	0.025***	0.021***
	(0.008)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)
Observations	94,329	232,788	233,177	234,416	234,799	235,233
Dep. var. mean (1993)	0.120	0.120	0.120	0.120	0.120	0.120
Number of provinces	15	15	15	15	15	15
Number of survey years	5	5	5	5	5	5
Number of districts	255	255	255	255	255	255
Province-Survey Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Urban FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Note: This Table presents the estimates from Equation 3. The dependent variable is a binary variable for migration, coded to one if the individuals move from their place of residence. I show the results by migration window, from 24 months to 6 months after an earthquake. Earthquakes are defined as earthquakes with an intensity of at least VII in some of their locations affected (Gignoux and Menéndez 2016). Observations are at the survey year level. Standard errors are clustered at the district level. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.1.22: Differences in characteristics between *induced-migrant* and *left-behind*

	Mean	Mean	Diff (2) - (1)
	Left-behind	Migrant	
Individual Characteristics			
age	0.091	-0.491	-0.575***
	(1.026)	(0.658)	(0.015)
non-muslim	0.003	-0.015	-0.043***
	(1.004)	(0.980)	(0.012)
non-javanese	-0.014	0.069	0.001
	(1.003)	(0.982)	(0.009)
married	0.069	-0.365	-0.421***
	(1.006)	(0.881)	(0.015)
at least primary	0.008	-0.041	-0.047***
	(0.999)	(1.002)	(0.015)
employed	0.041	-0.233	-0.267***
	(1.004)	(0.943)	(0.017)
siblings	-0.008	0.043	-0.020
	(0.985)	(1.080)	(0.014)
female siblings	-0.002	0.009	-0.022
	(1.009)	(0.947)	(0.015)
Parent Characteristics			
age father	0.034	-0.112	-0.144***
	(1.035)	(0.866)	(0.021)
at least primary father	-0.010	0.035	0.056***
	(1.000)	(0.999)	(0.019)
working father	0.002	-0.007	-0.035
	(1.003)	(0.991)	(0.072)
self-employed father	-0.051	0.274	0.361**
	(0.973)	(1.105)	(0.147)
in agriculture father	-0.020	0.068	-0.066***
	(0.993)	(1.021)	(0.025)
age mother	0.042	-0.144	-0.207***
	(1.042)	(0.825)	(0.019)
at least primary mother	-0.004	0.013	0.035*
	(1.001)	(0.997)	(0.018)
Household Characteristics			
arisan participation	-0.049	0.272	0.354***
	(0.990)	(1.011)	(0.014)
com. organization part	0.975	1.291	0.441***
	(1.578)	(1.866)	(0.023)
own house	0.030	-0.163	-0.142***
	(0.980)	(1.090)	(0.014)
own farm	-0.007	0.040	0.025*
	(0.994)	(1.032)	(0.015)
own livestock	-0.012	0.066	0.036**
	(0.985)	(1.073)	(0.014)
value own house	0.020	-0.106	-0.071***
	(1.040)	(0.738)	(0.015)
value farm	0.005	-0.029	-0.019
	(1.036)	(0.773)	(0.015)
savings	-0.003	0.019	0.034**
	(0.855)	(1.567)	(0.015)
labour income	0.014	-0.069	-0.017
	(1.028)	(0.844)	(0.020)
non-labour income	0.001	-0.005	0.006
	(0.989)	(1.056)	(0.016)
Observations	28,459	5,173	33,632

Note: This table shows along which dimensions *earthquake-induced migrant* and *left-behind* women differ. I report coefficient estimates together with 95% confidence intervals from a regression of an indicator variable for migrating right after an earthquake on socio-economic characteristics before an earthquake and urban fixed effects. *** p<0.01, ** p<0.05, * p<0.1

Table A.1.23: Conditional on same earthquake-exposure

	Mean	Mean	Diff (2) - (1)
	Left-behind	Migrants	
Economic Characteristics			
cement wall	0.035	-0.066	-0.003
	(0.987)	(1.020)	(0.015)
concrete roof	0.002	-0.036	0.004
	(1.009)	(0.819)	(0.016)
fixed assets	0.075	-0.019	-0.072***
	(0.913)	(1.020)	(0.024)
mobile assets	0.006	0.117	0.132***
	(0.991)	(0.816)	(0.026)
financial assets	0.020	-0.004	-0.065***
	(1.003)	(0.999)	(0.024)
own house	0.113	-0.196	-0.389***
	(0.916)	(1.105)	(0.017)
own farm	0.020	-0.067	-0.080***
	(1.016)	(0.941)	(0.016)
saving amount	0.003	0.007	-0.018
	(0.997)	(1.099)	(0.012)
rent stocks	0.002	-0.000	-0.000
	(1.037)	(0.986)	(0.000)
hh labour income	0.005	0.020	0.002
	(1.097)	(0.715)	(0.027)
hh non-labour income	0.013	-0.030	-0.047**
	(1.174)	(0.295)	(0.019)
Father Characteristics			
primary educ.	-0.022	0.048	0.140***
	(1.000)	(0.999)	(0.024)
non-educ	0.021	-0.045	-0.123***
	(1.008)	(0.981)	(0.025)
working	0.019	-0.045	-0.154
	(1.027)	(0.934)	(0.093)
self-employed	-0.078	0.233	0.422**
	(0.957)	(1.093)	(0.206)
public worker	0.066	-0.181	-0.180
	(1.015)	(0.943)	(0.226)
in agriculture	-0.068	0.130	0.210***
	(0.975)	(1.035)	(0.031)
Mother Characteristics			
age	0.033	-0.057	-0.124***
	(1.003)	(0.994)	(0.023)
primary educ.	-0.015	0.030	0.123***
	(1.003)	(0.994)	(0.023)
non-educ	0.015	-0.030	-0.123***
	(1.003)	(0.994)	(0.023)
married	-0.004	-0.001	0.022
	(1.005)	(1.001)	(0.023)
working	0.009	-0.007	-0.002
	(1.011)	(0.992)	(0.036)
Observations	23,913	8,159	33,632

Note: This table shows which dimensions *earthquake-induced migrant* and *left-behind* women differ, conditional on being equally exposed to an earthquake. I report coefficient estimates together with 95% confidence intervals from a regression of an indicator variable for migrating after an earthquake on socio-economic characteristics before an earthquake and urban fixed effects. *** p<0.01, ** p<0.05, * p<0.1

Table A.1.24: *Induced-migrant* vs *left-behind*, sanity checks for sibling counterfactual

	Below age 23			Below age 18		
	(1)	(2)	(3)	(4)	(5)	(6)
Earthquake	-0.009	-0.007	-0.010	-0.009	-0.004	-0.008
	(0.006)	(0.011)	(0.007)	(0.006)	(0.008)	(0.006)
*Earthquake * Migration*	0.053***	0.040***	0.041***	0.049**	0.056***	0.038***
	(0.011)	(0.014)	(0.010)	(0.024)	(0.015)	(0.010)
*Earthquake * Migration * Age Gap*	-0.001**					
	(0.001)					
*Earthquake * Migration * Age Gap * N. Female Sib*				-0.003		
				(0.007)		
Observations	188,616	74,712	113,904	188,616	98,640	160,548
Dep. var. mean	0.036	0.036	0.036	0.018	0.018	0.018
Number of provinces	15	15	15	15	15	15
Number of years	22	22	22	22	22	22
Number of districts	255	255	255	255	255	255

Note: This Table presents the estimates from Equation 4 where $Eq_{s,t} * Disp_{s,t}$ is the interaction of earthquakes with a migration right after an earthquake. The dependent variable is a binary variable for marriage, coded to one if the woman married at the age corresponding to the observation. Earthquakes are defined as earthquakes with an intensity of at least VII in some of its locations affected (Gignoux and Menéndez 2016). The sample is restricted to women with female siblings. Observations are at the level of a person's age at the month level (from 12 to 22 or the age of first marriage). The results show the estimates for girl-to-girl comparison within the same family. Columns (4), (5) and (6) perform the same analysis that Columns (1), (2) and (3) but for a sub-sample of ages from 12 to 17 (or age of first marriage). Standard errors are clustered at the district level. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.1.25: *Induced-migrant* vs *left-behind*, by destination

VARIABLES	getting married in district	get. married in province	get. married other province	get. married in district	get. married other IFLS prov.
	(1)	(2)	(3)	(4)	(5)
Earthquake	-0.018**	-0.013*	-0.016**	-0.017**	0.009
	(0.008)	(0.007)	(0.007)	(0.008)	(0.020)
*Earthquake * Migration*	0.037***	0.025*	0.003	0.028***	0.001
	(0.007)	(0.013)	(0.012)	(0.009)	(0.013)
*Earthquake * Migration * Area*				0.000*	
				(0.000)	
*Earthquake * Migration * Distance*					-0.000
					(0.000)
Observations	192,600	163,128	162,744	192,600	169,407

Note: This Table presents the estimates from Equation 4 where $Eq_{s,t} * Disp_{s,t}$ is the interaction of earthquakes with a migration right after an earthquake. The dependent variable is a binary variable for marriage, coded to one if the woman married at the age corresponding to the observation. Earthquakes are defined as earthquakes with an intensity of at least VII in some of its locations affected (Gignoux and Menéndez 2016). Observations are at the level of a person's age at the month level (from 12 to 22 or the age of first marriage). The results are with the province, urban at origin, age, birth year fixed effects, and covariates (religion and mother education for the year before the earthquake). I present the result by restricting the treated women who migrate induced by an earthquake to a sub-sample of *earthquake-induced migrant* women who migrate within a district (Column (1)), within the same province (Column (2)), and to other IFLS province (Column (3)). Column (4) use the same sub-sample of Column (1) with an interaction with the district area. Column (5) shows the results for *earthquake-induced migrant* who migrate outside their district, adding interaction to the distance between origin and destination. Standard errors are clustered at the district level. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.1.26: Place-based effects, *induced-migrant* vs *left-behind*

VARIABLES	getting married	getting married	getting married	getting married	getting married
	(1)	(2)	(3)	(4)	(5)
PANEL A: By Ethnicity Composition					
	within homeland in dest.	within/adjacent homeland in dest.	in homeland in origin no in dest.	in homeland in origin and in dest.	no homeland in origin but in dest.
Earthquake	-0.020***	-0.020***	-0.020***	-0.020***	-0.025***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)
*Earthquake * Migration*	0.032***	0.033***	0.027***	0.033***	0.031***
	(0.008)	(0.009)	(0.006)	(0.008)	(0.009)
*Earthquake * Migration* * Homeland	-0.008	-0.008	-0.004	-0.010	0.033
	(0.011)	(0.011)	(0.017)	(0.010)	(0.040)
Observations	227,088	227,088	227,088	227,088	75,660
PANEL B: By Development					
	population	pop. after eq	pop. diff.	night lights	night lights diff.
Earthquake	-0.015**	-0.019**	-0.010	-0.020***	-0.021***
	(0.007)	(0.008)	(0.009)	(0.007)	(0.007)
*Earthquake * Migration*	0.041***	0.035***	0.032***	0.039***	0.032***
	(0.007)	(0.006)	(0.008)	(0.008)	(0.006)
*Earthquake * Migration* * Development	-0.000***	0.000*	0.000**	-0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	170,340	164,052	131,508	224,892	224,364
PANEL C: By Marriage Market composition					
	unmarried pop in dest.	diff. unmarried pop orig. vs dest.	diff. unmarried pop orig. vs dest. aft eq.	eq. in dest.	destruc. in dest.
Earthquake	-0.016**	-0.019**	-0.011	-0.021***	-0.021***
	(0.007)	(0.008)	(0.009)	(0.007)	(0.007)
*Earthquake * Migration*	0.026***	0.032***	0.027***	0.009	0.026***
	(0.007)	(0.006)	(0.008)	(0.007)	(0.005)
*Earthquake * Migration* * Market	-0.000	-0.000	-0.000	0.028***	0.001***
	(0.000)	(0.000)	(0.000)	(0.009)	(0.000)
Observations	170,340	164,052	131,508	227,088	226,248

Note: This Table presents the estimates on how local marriage markets at the destination may affect results from Equation 4. The dependent variable is a binary variable for marriage, coded to one if the woman married at the age corresponding to the observation. Observations are at the level of person age at month level (from 12 to 22 or age of first marriage). The results are with the province, urban at origin, age, birth year fixed effects, and covariates (religion and mother education for the year before earthquake). I include an additional interaction on the market characteristics. In Panel A, I include different proxies of ethnicity composition: destination falls within their homeland (column 1), within or adjacent to their homeland (column 2), at origin in their homeland but not at the new destination (column 3), or origin and destination in their homeland (column 4), destination is within the homeland but not their origin (column 5). In Panel B, I include development proxies: population density (column 1), differences in population density between origin and destination after an earthquake (column 2), differences in population density between origin before an earthquake and destination after an earthquake (column 3), night light intensity at the destination after an earthquake (column 4), and differences in night light intensity between origin and destination after an earthquake (column 5). In Panel C, I evaluate how the marriage market composition at the destination: the sex ratio at the destination after an earthquake (column 1), differences between origin and destination after an earthquake (column 2), differences between origin before an earthquake and destination after an earthquake (column 3), an earthquake also hits the marriage market at the destination (column 4), number of houses destroyed at destination (column 5). Standard errors are clustered at district level. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.1.27: Counterfactual for earthquake-induced migrant women

	Below age 23			Below age 18		
VARIABLES	getting married	get. married	get. married	get. married	get. married	get. married
	(1)	(2)	(3)	(4)	(5)	(6)
PANEL A: Baseline Specification						
Earthquake	0.021***	0.000	-0.020***	0.017***	0.009***	0.007
	(0.004)	(0.004)	(0.007)	(0.003)	(0.003)	(0.007)
*Earthquake * Migration*	0.027***	0.026***	0.027***	-0.007	-0.008	-0.008
	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Observations	227,088	227,088	227,088	135,552	135,552	135,552
PANEL B: Later nduced-migrants						
Earthquake	0.030***	0.006	-0.019	0.010***	0.005*	-0.004
	(0.005)	(0.005)	(0.014)	(0.003)	(0.003)	(0.011)
Observations	73,464	73,464	73,464	44,004	44,004	44,004
PANEL C: Non-exposed to earthquakes						
Earthquake	0.020***	0.022***	0.021***	0.002	0.001	0.001
	(0.005)	(0.005)	(0.005)	(0.003)	(0.004)	(0.004)
Observations	432,192	432,192	432,192	258,684	258,576	258,684
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Urban FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth Year FE	No	Yes	Yes	No	Yes	Yes
Age FE	No	Yes	Yes	No	Yes	Yes
Controls	No	No	Yes	No	No	Yes

Note: This table displays the estimation results for the heterogeneity effect on the timing of marriage for *earthquake-induced migrant* women. This table includes three different counterfactuals. Panel A includes the baseline counterfactual: women exposed to earthquakes that didn't migrate right after an earthquake. Panel B reports the estimates for a counterfactual of latter *earthquake-induced migrant* women. And Panel C presents the results for non-exposed earthquakes counterfactual. The dataset is in a person-year panel format. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

D Mechanisms

Table A.1.28: Bride Price interaction, *induced-migrant* vs *left-behind*

	Below age 23			Below age 18		
	(1)	(2)	(3)	(4)	(5)	(6)
Earthquake	0.021***	0.001	-0.021***	0.017***	0.010***	0.009
	(0.004)	(0.004)	(0.007)	(0.003)	(0.003)	(0.008)
*Earthquake * Migration*	0.028***	0.028***	0.026***	-0.005	-0.005	-0.006
	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)
*Earthquake * Migration* * Brideprice	-0.007	-0.007	0.005	-0.012	-0.014	-0.011
	(0.010)	(0.010)	(0.006)	(0.009)	(0.009)	(0.008)
Observations	223,392	223,392	223,392	133,536	133,536	133,536
Dep. var. mean	0.036	0.036	0.036	0.018	0.018	0.018
Number of provinces	15	15	15	15	15	15
Number of years	22	22	22	22	22	22
Number of districts	255	255	255	255	255	255
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Urban FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth Year FE	No	Yes	Yes	No	Yes	Yes
Age FE	No	Yes	Yes	No	Yes	Yes
Controls	No	No	Yes	No	No	Yes

Note: This Table presents the estimates from Equation 4 where $Eq_{s,t} * Disp_{s,t}$ is the interaction of earthquakes with a migration right after an earthquake. The dependent variable is a binary variable for marriage, coded to one if the woman married at the age corresponding to the observation. Earthquakes are defined as earthquakes with an intensity of at least VII in some of its locations affected (Gignoux and Menéndez 2016). Observations are at the level of the person's age at the month level (from 12 to 22 or the age of first marriage). Column (1) presents the results without age, birth year fixed effects, and covariates. Column (2) includes age and birth year fixed effects. Column (3) controls for baseline characteristics (religion and mother education for the year before the earthquake). Columns (4), (5) and (6) perform the same analysis that Columns (1), (2) and (3) but for a sub-sample of ages from 12 to 17 (or age of first marriage). Standard errors are clustered at the district level. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.1.29: Bride Price and Matrilocal relationship, *induced-migrant* vs *left-behind*

VARIABLES	All getting married	Matrilocal getting married	Non-Matrilocal getting married
	(1)	(2)	(3)
Earthquake	-0.019**	0.099	-0.013
	(0.009)	(0.252)	(0.011)
*Earthquake * Migration*	0.037***	0.165	0.034***
	(0.008)	(0.127)	(0.008)
Observations	46,788	1,572	42,456
Dep. var. mean	0.036	0.036	0.036
Number of provinces	15	15	15
Number of years	22	22	22
Number of districts	255	255	255
Province FE	Yes	Yes	Yes
Urban FE	Yes	Yes	Yes
Age FE	No	Yes	Yes
Birth Year FE	Yes	Yes	Yes
Controls	No	No	Yes

Note: This Table presents the estimates from Equation 4 where $Eq_{s,t} * Disp_{s,t}$ is the interaction of earthquakes with a migration right after an earthquake. The dependent variable is a binary variable for marriage, coded to one if the woman married at the age corresponding to the observation. Earthquakes are defined as earthquakes with an intensity of at least VII in some of its locations affected (Gignoux and Menéndez 2016). The sample is restricted to women engaged in the bride price tradition. Observations are at the level of a person's age at the month level (from 12 to 22 or the age of first marriage). The estimation includes province, age and birth year fixed effects and controls for baseline characteristics (religion and mother education for the year before the earthquake). Column (1) presents the entire sample. Column (2) includes only a sub-sample of matrilocal women (husband joins wife's household). Column (3) reports the estimates for a sub-sample of non-matrilocal women (wife joins their husband's household or creates their own household). Standard errors are clustered at the district level. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1. I do not report the estimates restricting to child marriage, because the matrilocal sample is small.

Table A.1.30: Matrilocality interaction, *induced-migrant* vs *left-behind*

VARIABLES	getting married	getting married	getting married
	(1)	(2)	(3)
Earthquake	0.021***	0.001	-0.021***
	(0.004)	(0.004)	(0.007)
*Earthquake * Migration*	0.027***	0.026***	0.027***
	(0.006)	(0.005)	(0.005)
*Earthquake * Migration* * Matrilocal	0.515***	0.517***	0.537***
	(0.145)	(0.145)	(0.140)
Observations	225,108	225,108	225,108
Dep. var. mean	0.036	0.036	0.036
Number of provinces	15	15	15
Number of years	22	22	22
Number of districts	255	255	255
Province FE	Yes	Yes	Yes
Urban FE	Yes	Yes	Yes
Age FE	No	Yes	Yes
Birth Year FE	No	Yes	Yes
Controls	No	No	Yes

Note: This Table presents the estimates from Equation 4 where $Eq_{s,t} * Disp_{s,t}$ is the interaction of earthquakes with a migration right after an earthquake. The dependent variable is a binary variable for marriage, coded to one if the woman married at the age corresponding to the observation. Earthquakes are defined as earthquakes with an intensity of at least VII in some of its locations affected (Gignoux and Menéndez 2016). Observations are at the level of the person's age at the month level (from 12 to 22 or the age of first marriage). The estimation includes age and birth year fixed effects and controls for baseline characteristics (religion and mother education for the year before the earthquake). Column (1) presents the estimates without age, year-of-birth fixed effects and covariates. Column (2) includes age and year-of-birth fixed effects. In Column (3) add the covariates. Standard errors are clustered at the district level. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1. I do not report the estimates restricting to child marriage, because the matrilocal sample is small.

Table A.1.31: Inter-ethnic Marriage, *induced-migrant* vs *left-behind*

	Below age 23			Below age 18		
VARIABLES	getting married	get. married	get. married	get. married	get. married	get. married
	(1)	(2)	(3)	(4)	(5)	(6)
Earthquake	-0.000	-0.001	-0.001*	0.000	0.000	0.000
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
*Earthquake * Migration*	0.001	0.001	0.001	0.000	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)
Observations	227,088	227,088	227,088	135,552	135,552	135,552
Dep. var. mean	0.036	0.036	0.036	0.018	0.018	0.018
Number of provinces	15	15	15	15	15	15
Number of years	22	22	22	22	22	22
Number of districts	255	255	255	255	255	255
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Urban FE	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	No	Yes	Yes	No	Yes	Yes
Birth Year FE	No	Yes	Yes	No	Yes	Yes
Controls	No	No	Yes	No	No	Yes

Note: This Table presents the estimates from Equation 4 where $Eq_{s,t} * Disp_{s,t}$ is the interaction of earthquakes with a migration right after an earthquake. The dependent variable is a binary variable for marriage, coded to one if the woman married at the age corresponding to the observation. Earthquakes are defined as earthquakes with an intensity of at least VII in some of its locations affected (Gignoux and Menéndez 2016). Observations are at the level of the person's age at the month level (from 12 to 22 or the age of first marriage). Column (1) presents the results without age, birth year fixed effects, and covariates. Column (2) includes age and birth year fixed effects. Column (3) controls for baseline characteristics (religion and mother education for the year before the earthquake). Columns (4), (5) and (6) perform the same analysis that Columns (1), (2) and (3) but for a sub-sample of ages from 12 to 17 (or age of first marriage). Standard errors are clustered at the district level. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.1.32: Transfer at the moment of the marriage, *induced-migrant* vs *left-behind*

	Full sample			Bride Price subsample		
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	bride price value	bp value	bp value	bp value	bp value	bp value
	PANEL A: Contemporary price- All					
Earthquake	-0.031	-0.032	-0.013	0.119	0.124	0.265*
	(0.060)	(0.061)	(0.070)	(0.142)	(0.140)	(0.149)
Observations	226,620	217,200	166,704	70,764	67,464	52,836
	PANEL B: Contemporary price- Bellow 18					
Earthquake	-0.002	-0.001	-0.003	0.032	0.041	0.008
	(0.043)	(0.044)	(0.047)	(0.146)	(0.150)	(0.152)
Observations	145,920	137,844	107,364	46,332	43,500	34,656
Province/ Urban FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth Yr/ Age FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	No	Yes	Yes
	PANEL C: Decomposition by groom					
	all	mover	native	all	mover	native
Earthquake	-0.109	0.195	0.039	-0.428	-0.318	0.192
	(0.146)	(0.203)	(0.192)	(0.513)	(0.765)	(0.474)
Observations	24,480	3,732	20,748	9,276	1,032	8,244

Note: This Table shows the estimates on transfer at the moment of the marriage. The dependent variable is a continuous variable for payment at marriage (*bride payment*) at the age of marriage. Earthquakes are defined as earthquakes with an intensity of at least VII in some of its locations affected (Gignoux and Menéndez 2016). Observations are at the women's level. Columns (1), (2), and (3) show the estimates for the entire sample. In columns (4), (5), and (6), I restrict the sample to women traditionally engaged in bride price. Columns (1) and (4) present the results without covariates. Columns (2) and (5) includes covariates (religion and mother's education before an earthquake) and woman's education fixed effects. Columns (3) and (6) controls also for the spouse's age at marriage. Panel A reports the results when marriage is after 17. Panel B presents the same analysis but for marriages below 17. Panel C shows the results by sub-sample of spouse's origin: *earthquake-induced migrant* or *native*. *Earthquake-induced migrant* are men that suffered an earthquake and migrated right after an earthquake. I define *native* as grooms that are not classified as *earthquake-induced migrant*. The number of observations decreases because not every woman has data on her spouse's origin. Standard errors are clustered at the district level. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.1.33: Groom's education at marriage, *induced-migrant* vs *left-behind*

	Below age 23			Below age 18		
	all	matrilocal	non-matrilocal	all	matrilocal	non-matrilocal
VARIABLES	educ. spouse	educ. spouse	educ. spouse	educ. spouse	educ. spouse	educ. spouse
	(1)	(2)	(3)	(4)	(5)	(6)
Earthquake	-0.008	-0.197	-0.040	-0.055	-0.255	-0.092
	(0.088)	(0.363)	(0.091)	(0.098)	(0.358)	(0.102)
Observations	43,848	3,828	39,612	34,908	3,132	31,392
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Urban FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Note: This Table shows the estimates on spouse's education at marriage by matrilocal norms. The dependent variable is a continuous variable for the education gap between spouses at marriage. Earthquakes are defined as earthquakes with an intensity of at least VII in some of its locations affected (Gignoux and Menéndez 2016). Observations are at the woman level. The estimation includes province, urban residence at origin, age, year of birth, woman's education fixed effects, and covariate (religion). Column (1) includes the entire sample. Column (2) restricts the sample to matrilocal women, and Column (3) to non-matrilocal women. Columns (4), (5) and (6) perform the same analysis that Columns (1), (2) and (3) but for marriages below 17. Standard errors are clustered at district level. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.1.34: Heterogeneity effect of earthquakes, by other cultural norms

	Below age 23			Below age 18		
VARIABLES	getting married	get. married	get. married	get. married	get. married	get. married
	(1)	(2)	(3)	(4)	(5)	(6)
	Polygyny					
	all	polygyny	non-polygyny	all	polygyny	non-polygyny
Earthquake	-0.020***	-0.012	-0.029***	0.007	-0.001	0.006
	(0.007)	(0.012)	(0.008)	(0.007)	(0.006)	(0.009)
*Earthquake * Migration*	0.027***	0.021*	0.028***	-0.008	-0.002	-0.009
	(0.005)	(0.011)	(0.006)	(0.005)	(0.011)	(0.006)
Observations	227,088	33,468	191,640	135,552	19,308	115,164
	Matrilineality					
	all	matri.	non-matri.	all	matri.	non-matri.
Earthquake	-0.020***	-0.001	0.030	0.007	0.009	0.007
	(0.007)	(0.012)	(0.023)	(0.007)	(0.010)	(0.007)
*Earthquake * Migration*	0.027***	0.016	0.059***	-0.008	-0.002	-0.075*
	(0.005)	(0.010)	(0.017)	(0.005)	(0.013)	(0.032)
Observations	227,088	25,056	17,928	135,552	14,424	10,704
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Urban FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table displays the estimation results for the effect of earthquakes on the timing of marriage between *induced-migrant* and *left-behind* women (equation (4)) by cultural norms. Polygyny is when a man has more than one wife. Matrilineality refers to the organization of family relationships in societies according to lines of descent from female ancestors. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.1.35: Contemporary engagement in bride price and matrilocal traditions

VARIABLES	getting married	getting married	getting married
	(1)	(2)	(3)
	PANEL A: Bride Price tradition		
	All sample	Bride Price	Non-Bride Price
Earthquake	-0.020***	0.014	0.005
	(0.007)	(0.047)	(0.003)
*Earthquake * Migration*	0.027***	-0.001	0.004
	(0.005)	(0.036)	(0.004)
Observations	227,088	11,220	35,568
	PANEL B: Matrilocality tradition		
	All sample	Matrilocal	Non-Matrilocal
Earthquake	-0.020***	-0.021***	0.000
	(0.007)	(0.006)	(0.000)
*Earthquake * Migration*	0.027***	0.021***	0.000
	(0.005)	(0.005)	(0.000)
Observations	227,088	225,984	1,104
Dep. var. mean	0.036	0.036	0.036
Observations	585,816	585,816	585,816
Number of provinces	15	15	15
Number of years	22	22	22
Number of districts	255	255	255
Province FE	Yes	Yes	Yes
Urban FE	Yes	Yes	Yes
Birth Year FE	No	Yes	Yes
Age FE	No	Yes	Yes
Controls	No	No	Yes

Note: This Table presents the estimates from Equation 4 by marriage norms contemporary engagement: *bride price* and *matrilocality* traditions. *Bride price* tradition is a payment from the groom (or groom's family) to the bride (or bride's family) at the moment of the marriage. In Indonesia, doesn't exist a payment from the bride to the groom's family (*dowry*). *Matrilocality* tradition is whereby the husband joins the wife's household after the marriage. When the wife joins the husband's household or settles down in a new household is known as *patrilocality* or *neolocality.* The dependent variable is a binary variable for marriage, coded to one if the woman married at the age corresponding to the observation. Estimates include province, urban, age and birth year fixed effects and control for baseline characteristics (religion and mother education for the year before the earthquake). Observations are at the level of a person's age at the month level (from 12 to 22 or the age of first marriage). Panel A reports the results by *bride price* sub-sample. Panel B reports the results by *matrilocal* women sub-sample. Standard errors are clustered at the district level. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Child marriage estimates are non-significant.

Table A.1.36: Earthquake effects on timing of marriage, by population outflow

	Below age 23			Below age 18		
VARIABLES	getting married	get. married	get. married	get. married	get. married	get. married
	(1)	(2)	(3)	(4)	(5)	(6)
PANEL A: Population change at district level						
Earthquake	0.039***	0.018**	0.018***	0.027***	0.019***	0.019***
	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)
Earthquake * Outflow	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	131,844	131,844	131,844	78,072	78,072	78,072
PANEL B: Change in Unmarried Population below 23 Sex ration at district level						
Earthquake	0.026***	0.004	0.004	0.011***	0.003	0.003
	(0.006)	(0.006)	(0.006)	(0.004)	(0.004)	(0.004)
Earthquake * Sex Ratio change	-0.000	-0.000	-0.000	-0.000*	-0.000*	-0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	131,844	131,844	131,844	78,072	78,072	78,072
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Urban FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth Year FE	No	Yes	Yes	No	Yes	Yes
Age FE	No	Yes	Yes	No	Yes	Yes
Controls	No	No	Yes	No	No	Yes

Note: This Table presents the earthquake results on the dependent variable by population outflow: annual marriage hazard. The dependent variable is a binary variable for marriage, coded to one if the woman married at the age corresponding to the observation. Earthquakes are defined as earthquakes with an intensity of at least VII in some of its locations affected (Gignoux and Menéndez 2016). Observations are at the level of a person's age at the month level (from 12 to 22 or the age of first marriage). I measure population outflow using population data at the district level from the Indonesian Population Census (1990, 2000, and 2010). The baseline specification is presented in Equation 1. Column (1) presents the results without age, birth year fixed effects, and covariates. Column (2) includes age and birth year fixed effects. Column (3) controls for baseline characteristics (religion and mother education for the year before the earthquake). Columns (4), (5) and (6) perform the same analysis that Columns (1), (2) and (3) but for a sub-sample of ages from 12 to 17 (or age of first marriage). Standard errors are clustered at the district level. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.1.37: Effect of earthquakes on education

	Below age 23			Below age 18		
VARIABLES	attending sch.	attending sch.	attending sch.	attending sch.	attending sch.	attending sch.
	(1)	(2)	(3)	(4)	(5)	(6)
PANEL A: School attendance						
Earthquake	-0.026*	-0.023	-0.018	-0.049***	-0.040***	-0.034***
	(0.015)	(0.016)	(0.014)	(0.015)	(0.015)	(0.013)
Observations	494,988	494,988	494,988	301,284	301,284	301,284
PANEL B: Educational attainment						
Earthquake	0.043	0.001	0.005	0.021	-0.025	-0.018
	(0.035)	(0.026)	(0.022)	(0.022)	(0.021)	(0.018)
Observations	539,028	539,028	539,028	315,168	315,168	315,168
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Urban FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth Year FE	No	Yes	Yes	No	Yes	Yes
Age FE	No	Yes	Yes	No	Yes	Yes
Controls	No	No	Yes	No	No	Yes

Note: This Table presents the earthquake results on school attendance and educational attainment. Earthquakes are defined as earthquakes with an intensity of at least VII in some of its locations affected (Gignoux and Menéndez 2016). Observations are at the level of a person's age at the month level (from 12 to 22 or the age of first marriage). The baseline specification is presented in Equation 1. Column (1) presents the results without age, birth year fixed effects, and covariates. Column (2) includes age and birth year fixed effects. Column (3) controls for baseline characteristics (religion and mother education for the year before the earthquake). Columns (4), (5) and (6) perform the same analysis that Columns (1), (2) and (3) but for a sub-sample of ages from 12 to 17 (or age of first marriage). Panel A show the results with an outcome on attending school. The outcome is defined as one if the woman i is attending the school at age a, zero otherwise. Panel B shows the results on the level of educational attainment. Standard errors are clustered at the district level. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

E Other tests

Table A.1.38: Sample definition: Women at least 25 years old

	Below age 23			Below age 18		
VARIABLES	getting married	get. married	get. married	get. married	get. married	get. married
	(1)	(2)	(3)	(4)	(5)	(6)
	PANEL A: Earthquakes effects					
Earthquake	0.009***	0.007**	0.006**	0.011***	0.010***	0.009**
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
Observations	45,914	45,914	45,914	24,627	24,627	24,627
	PANEL B: Heterogeneity effects between mover and non-mover women					
Earthquake	0.013***	0.001	-0.019**	0.017***	0.013***	0.011
	(0.004)	(0.004)	(0.008)	(0.004)	(0.004)	(0.009)
*Earthquake * Migration*	0.048***	0.048***	0.049***	-0.007	-0.009	-0.009
	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)
Observations	19,551	19,551	19,551	10,470	10,470	10,470
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Urban FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth Year FE	No	Yes	Yes	No	Yes	Yes
Age FE	No	Yes	Yes	No	Yes	Yes
Controls	No	No	Yes	No	No	Yes

Note: This Table presents the estimation results of Tables 1 and 3 for a new sample. I restrict the sample to young women at least 25 years old at the last interview. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.1.39: Placebo groups for mover women

	Below age 23			Below age 18		
VARIABLES	getting married	get. married	get. married	get. married	get. married	get. married
	(1)	(2)	(3)	(4)	(5)	(6)
	PANEL A: Women exposed to Earthquakes at age 23 or older					
Earthquake	0.020*	-0.003	-0.013	0.019**	0.003	0.012
	(0.010)	(0.011)	(0.013)	(0.009)	(0.008)	(0.012)
*Earthquake * Migration*	-0.002	0.007	0.007	-0.002	-0.001	-0.000
	(0.013)	(0.013)	(0.012)	(0.015)	(0.015)	(0.015)
Observations	40,224	40,224	40,224	24,540	24,540	24,540
	PANEL B: Effects of voluntary migration on Marriage					
Voluntary Migration	0.502***	-0.208**	-0.271***	-0.176	-0.383***	-0.434***
	(0.098)	(0.092)	(0.091)	(0.133)	(0.132)	(0.130)
Observations	61,379	61,379	61,379	38,439	38,439	38,439
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Urban FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth Year FE	No	Yes	Yes	No	Yes	Yes
Age FE	No	Yes	Yes	No	Yes	Yes
Controls	No	No	Yes	No	No	Yes

Note: This Table presents the estimation results of Table 3 for two placebo analyses. Panel A shows the results for the exposure to earthquakes at age 23 or older. Panel B shows the effects of voluntary migration on the timing of marriage. Voluntary migration is defined as every migration excluding the migration called *forced migration*. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.1.40: Men sample

	Below age 23			Below age 18		
VARIABLES	getting married	get. married	get. married	get. married	get. married	get. married
	(1)	(2)	(3)	(4)	(5)	(6)
PANEL A: Earthquake Effects						
Earthquake	0.001	0.001	0.001	0.001	0.000	0.000
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	54,703	54,703	54,703	30,441	30,441	30,441
PANEL B: Heterogeneity effects between mover and non-mover men						
Earthquake	0.008***	-0.000	-0.002	0.002*	0.001	0.002
	(0.002)	(0.001)	(0.003)	(0.001)	(0.001)	(0.002)
*Earthquake * Migration*	0.008**	0.007**	0.008**	-0.001	-0.001	-0.001
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Observations	21,293	21,293	21,293	11,830	11,830	11,830
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Urban FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth Year FE	No	Yes	Yes	No	Yes	Yes
Age FE	No	Yes	Yes	No	Yes	Yes
Controls	No	No	Yes	No	No	Yes
PANEL C: Bride Price Heterogeneity						
	All sample	Bride Price	Non-Bride Price	All sample	Bride Price	Non-Bride Price
Earthquake	-0.002	0.005	-0.003	0.002	0.008**	-0.000
	(0.003)	(0.006)	(0.006)	(0.002)	(0.004)	(0.002)
*Earthquake * Migration*	0.008**	0.006	0.009**	-0.001	0.001	-0.002
	(0.003)	(0.006)	(0.004)	(0.002)	(0.004)	(0.002)
Observations	21,293	4,784	15,964	11,830	2,650	8,880

Note: This Table presents the results for Tables 1, 3 and 4 for a sample of men. The sample includes all men at least 23 in the last interview and born after 1980. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

F Additional analysis

F.1 Mechanisms: *earthquake-induced migrants* versus *left-behind women*

F.1.1 Theoretical Framework

In this section, I extend a model, originally developed in (Corno, Hildebrandt, and Voena 2020)), to study how forced displacement can affect the timing of marriage of displaced women. This model contributes to the literature in threefold aspects. First, it looks at a destructive income shock and its resulting induced migration. Second, migration implies a change in the marriage market from the origin to the new destination. Third, the new marriage market is characterized by the lack of local social networks for the new migrants. Furthermore, in theory, the assumption is that they move to a marriage market where *native* 's credit capacity is not affected by the income shock.

F.1.1.1 Setup There is a unit mass of households with a daughter and a unit mass of households with a son. There are two periods, which correspond to two life stages, childhood (t=*1*) and adulthood (t=*2*). Since men typically marry younger women, period one may correspond to childhood for a woman and frequently to young adulthood for a man.

Households obtain a payoff from consumption in each period, where the payoff function ($u(.)$) is increasing and concave in consumption. For simplicity, assume that $u(.) = log(.)$. Future payoffs are discounted by a factor δ.

In each period, household income depends on three components: (1) a permanent income component (y_t) that is an independent and identical draw from a continuous distribution; (2) an idiosyncratic income component (ϵ_t) that is an independent and identical draw from a uniform distribution over [0,1]; and (3) the contribution of adult children labour (w^m,w^f), conditional on adult children remain in the household. I assume that $w_1^f = 0$. Forced displacement is a negative income shock for forcibly displaced households. I capture this shock by allowing the household's permanent income to reduce by a fraction (d) that goes from 0 to 1, where d=0 in non-displaced households. I assume that children migrate together with their parents. Therefore, in period t, the total income of a household i with an adult offspring is equal to y_t(1-d_t) + ϵ_t + $w^{m,f}$.

The daughter's family received a bride payment (p_t) from the groom's family at the time of their marriage. I consider households to be matrilocal: upon marriage men move to the bride's family and contribute to its budget in w_m>0. As a consequence, in the daughter's family $w_2^f > 0$ and $w_2^m > 0$. This framework looks at marriage markets at the destination, not at the origin. Therefore, being a displaced household implies settling down in a new marriage market, and, with their offspring's marriage, households acquire new socioeconomic networks at the destination, which deliver utility $\eta_t \geq 0$. A potential utility gain of a woman's family stems from marrying off a daughter (for example, not experiencing the stigma associated with non-married women), denoted as $\xi^f \geq 0$.

My framework assumes that every individual in the cohort gets married in childhood or adulthood. With this assumption, an income shock from displacement may potentially affect the timing of marriage, but will not affect the probability of marriage, as everyone is married in adulthood.

F.1.1.2 Adulthood In adulthood, marriage occurs only if the payoff from marriage is larger than when their offspring is unmarried.

Supply and Demand for Brides

$$Supply : \ln(y_2(1-d_2)+\epsilon_2^w+w^f+p_2+w_2^m)+\eta_2^f+\xi^f > \ln(y_2(1-d_2)+\epsilon_2^w+w^f) \tag{1.6}$$

$$Demand : \ln(y_2(1-d_2)+\epsilon_2^m-w_2^m-p_2)+\eta_2^m > \ln(y_2(1-d_2)+\epsilon_2^m+w_2^m) \tag{1.7}$$

Displaced brides are demanded by displaced or *native* households with sons. I assume that η_2 and d_2 are equal to 0 in the demand by *native* households.

Equilibrium bride price in Adulthood The conditions of equations (1), (2) and (3) imply that there are two equilibrium bride prices in adulthood: one for each demand. These conditions give a lower bound on the equilibrium: $p_2^* \geq \frac{(1-exp(\eta_2^f+\xi^f))}{exp(\eta_2^f+\xi^f)}$ $(y_2(1-d_2)$ $+$ ϵ_2^w $+$ $w^f)$ - w_2^m, which implies that bride price must be at least as much as the lower bound.

The upper bound on the equilibrium bride price in adulthood is equal to $p_2^* \leq \frac{(exp(\eta_2^m)-1)}{exp(\eta_2^m)}$ $(y_2(1-d_2)$ $+$ $\epsilon_2^m)$ - $\frac{(exp(\eta_2^m)+1)}{exp(\eta_2^m)}$ w_2^m for displaced households, and $p_2^* \leq$ -2 w_2^m for *native* households. A simple example of equilibrium is when men make a take-it-or-leave-it offer to the woman's parents, and the parents decide whether or not to accept.

Given the payment p_2^*, the lower bound decreases with η_2^f and w_2^m, and the upper bound decreases with w_2^m and increases, in the case of displaced households, with η_2^m. The intuition of this result is that daughter's family would decrease their received payment if their socioeconomic network utility gain or the contribution of the groom in the labour market increases. On the other hand, the son's families decrease their payment if their son's labour return to the bride's household increases. And households with displaced sons are willing to increase their payment as a trade-off for a higher network utility gain. In what follows, I assume that there exists a payment $p_2^* \in [\underline{p_2}, \overline{p_2}]$ that satisfies these conditions.

Proposition 1. *There exists a non-empty interval $[\underline{p_2}, \overline{p_2}]$ such that, with marriage transfer $p_2^* \in [\underline{p_2}, \overline{p_2}]$, everyone who is single at the beginning of the second period marries, as long as the gains from marriage for women, η_2^f and ξ^f, are sufficiently large.* Proof. See Appendix D

F.2.1.3 Childhood A household with a child will marry its child in childhood if and only if the household's payoff from a marriage in childhood is greater than in adulthood, that is:

Supply of Child Brides: Households with a daughther

$$\ln(y_1(1-d)+\epsilon_1^w+p_1+w_1^m)+\eta_1^f-[\ln(y_1(1-d)+\epsilon_1^w)] >$$

$$\delta[E[\ln(y_2+\epsilon_2^w+w^f+p_2+w_2^m)+\eta_2^f+\xi^f] \quad -$$

$$[E[\ln(y_2+\epsilon_2^w+w^f)]] \qquad (1.8)$$

A marginal household with a daughter is the one that has an idiosyncratic income realization (ϵ^w) such that it is indifferent between marrying her in childhood and marrying her in adulthood. In households with first-period income realizations lower than the threshold, ϵ_1^{w*} ($\epsilon_1^w \geq \epsilon_1^{w*}$), parents will want to marry their daughters in childhood. Since idiosyncratic incomes are uniformly distributed over the support [0,1], the mass of child daughters in the marriage market is ϵ_1^{w*}. Define the right hand-side term as $\Omega_f = \delta$ [E[ln $(y_2 + \epsilon_2^w + w^f + p_2 + w_2^m) + \eta_2 + \xi^f]$ - $[E[\ln$ (y2+ $\epsilon_2^w + w^f)]]$. Hence, the supply of child brides is given by:

$$SS_{brides} = \frac{y_1(1-d)(\eta_1^f-1)-(p_1+w_1^m)}{1-\eta_1^f} \qquad (1.9)$$

where $\eta_1^f = exp(\Omega_f - \eta_1^f)$. Thus, the supply of child brides is decreasing in network utility η_1^f and in households' permanent income (y_1). Therefore it is increasing in displacement (d). However, how supply is affected by bride price (p_1), and groom's labor contribution (w_1^m) depend on the value of η_1^f. If η_1^f >1, supply is increasing in both bride price (p_1) and groom's labour contribution (w_1^m). However, if η_1^m <1, we observe the opposite direction.

Demand for Child Brides: Households with a son

$$\ln(y_1(1-d)+\epsilon_1^m - w_1^m - p_1)+\eta_1 - [\ln(y_1(1-d)+\epsilon_1^m + w_1^m)] >$$

$$\delta[E[\ln(y_2+\epsilon_2^m - w_2^m - p_2)+\eta_2]] -$$

$$[E[\ln(y_2+\epsilon_2^m + w_2^m)]] \qquad (1.10)$$

where I assume that d and η_t are equal to 0 in the demand by *native* households.

For $\epsilon_1^m \geq \epsilon_1^{m*}$, men want to also marry in the first period. Hence, because of the uniform assumption, a measure 1 - ϵ_1^{m*} wants to get married. Define the right handside term as $\Omega_m = \delta$ [E[$\ln(y_2+\epsilon_2^m - w_2^m - p_2)+\eta_2]] - [E[\ln(y_2+\epsilon_2^m + w_2^m)]]]$. The demand for brides, again defined on the [0,1] interval, takes the form:

$$DD_{brides} = 1 - [\frac{y_1(1-d)(\eta_1^m - 1) + w_1^m(\eta_1^m + 1) + p_1]}{1-\eta_1^m}] \qquad (1.11)$$

where $\eta_1^m = exp(\Omega_m - \eta_1^m)$. The demand is increasing in network utility η_1^m and in households' permanent income (y_1). As a result, demand is decreasing in displacement cost (d). However, how demand is affected by bride price (p_1), and groom's labor contribution (w_1^m) depend on the value of η_1^m. If η_1^m >1, demand is increasing in both bride price (p_1) and groom's labour contribution (w_1^m). However, if η_1^m <1, we observe the opposite direction.

Equilibrium bride price and quantity in the marriage market Equilibrium marriage payment which clears the marriage market in the first period, is the one that solves *D(*y_1*,* p_1^**)* = *S(*y_1*,* p_1^**)*.

$$p_1^* = \frac{y_1(1-d)[(1-\eta_1^f)(\eta_1^m - 1) + (1-\eta_1^m)(\eta_1^f - 1)] + 2w_1^m - (1-\eta_1^f)(1-\eta_1^m)}{2-\eta_1^m - \eta_1^f} \qquad (1.12)$$

where η_1^f >0 and η_1^m ≥0. This implies that the bride price in equilibrium is increasing in the level of income if $\eta_1^f + \eta_1^m < 2$. The price is increasing or decreasing in the groom's labour contribution (w_1^m) depending if $\eta_1^f + \eta_1^m <$ or $>$ than 2, respectively. The relationship between bride price in equilibrium and network utility (η_1^f and η_1^m) depends on the value of $\eta_1^f + \eta_1^m$. If, $\eta_1^f + \eta_1^m > 2$, the price is decreasing. And, it is increasing if $\eta_1^f + \eta_1^m < 2$.[45]

[45] When η_1^m=0, η_1^f need to be > 1. Otherwise, the price decrease in female network utility (η_1^f).

Equilibrium quantities are estimated by substituting the equilibrium price in the supply or demand equation. Equilibrium quantities of child marriages are equal to

$$Q^*_{y_1} = \frac{y_1(1-d)[\phi_1(\eta_1^f - 1) - \phi_2^2[\phi_2(\eta_1^m - 1) + (1-\eta_1^m)(\eta_1^f - 1)]] - w_1^m[\phi_1 + 2\phi_2^2] + \phi_2^3(1-\eta_1^m)}{\phi_1\phi_2} \tag{1.13}$$

where $\phi_1{=}2-\eta_1^m-\eta_1^f$ and $\phi_2{=}\ 1-\eta_1^f$

Proposition 2. *Marriage payments are affected by income-level. However, the direction of the effects depends on the value of the network utility that the displaced bride and displaced groom gain from marriage.*

Proposition 3. *Income decreases the number of child marriages in equilibrium, as long as network gains from marriage, η_1^f and η_1^m, are sufficiently low.*

Proposition 4. *How groom's labour contribution affects the number of child marriages in equilibrium depends on how large or low network gains from marriage, η_1^f and η_1^m, are.*

Proofs for each proposition are in appendix J.

F.2 Other mechanisms for induced-migrant women

In what follows, I examine whether my findings may be affected by different local characteristics or different behaviour of the displaced population after displacement upon their arrival at their destinations. Appendix C.3 provides additional details.

F.2.1 Differential fertility. To evaluate the length of time (duration) that adult women spend without being pregnant after an earthquake, I generate a new sample of women in their fertility age. Namely, to avoid including never-fertile women in the sample, my new sample is restricted to women between 15 and 28 when they were interviewed for the first time. Furthermore, I restrict my sample to those women with at least two observations over time.

Moreover, I convert the data into a person-year panel format. Hence, a woman contributes 22 observations to the sample, one per year between 1993 and 2014. I merge these individual data with earthquake data and covariates at the year level.

This analysis examines the heterogeneity effects between *earthquake-induced migrants* versus *left-behind women.* I restrict the analysis to women exposed to earthquakes and compare displaced to stayers women. I estimate the following specification:

$$Y_{i,s,k,a} = \beta_0 + \beta_1 Eq_{s,p,a,m} + \beta_2 Eq_{s,p,a,m} * Disp_{i,s,a} + Eq_{s,p,a,m} * X_i + \alpha_p + \gamma_a + \delta_k + \zeta_{u_o} + \epsilon_{i,d} \tag{1.14}$$

where $Y_{i,s,k,a}$ is a binary variable coded as 1 at the age the woman gets pregnant and zero otherwise. The exposure to a destructive earthquake, $Eq_{s,p,a,m}$, switches to one from the occurrence of an earthquake in the sub-district of resident s at the age a and month m, zero otherwise. $Disp_{i,s,a}$ switches to one if displaced after the earthquake $Eq_{s,p,a,m}$. I control for province α_p, age, γ_a, cohort of birth, δ_k and urban at origin, ζ_{u_o}, fixed-effects. I further control for covariates, X_i (being married and employed). Standard errors are clustered at the district level.

Columns (1) to (4) in Table A.1.41 study the possibility that differential fertility may affect my results. For example, *earthquake-induced migrant women* may have chosen higher fertility to increase the future labour force within their household. More offspring could then help compensate for the economic shock from displacement in the long-term. An additional child may be translated into a substitution effect of their young daughter for the newborn. This hypothesis may be particularly true if we observe a preference shift for sons instead of daughters.

Therefore, I first wonder if there is an increase in fertility preferences, and secondly, if this change in fertility preferences is translated into a current increase in the number of children, to end up observing if displacement increases the son preferences. I find that *earthquake-induced migrant women* between ages 15 and 49 are 2.8 percentage points (pp) more likely to be pregnant in the same year (Column (1)). And, *earthquake-induced migrant women* are 28% more likely to have an additional pregnancy (Column (2)) and 21% to have an additional child (Column (3)). It seems that *earthquake-induced migrant women* are three percentage points (pp) more likely to prefer to have a future son than a daughter (Column (6)). These results speak in favour of fertility preferences as a potential channel of my results.

Table A.1.41: Differential Fertility Preferences and Son preferences

VARIABLES	(1) being pregnant.	(2) n. pregnancies.	(3) n. children	(4) annual vari. n. children	(5) daughter preferences	(6) pref. son
*Earthquake * Migration*	0.028***	0.287***	0.212***	0.032***	-0.029	0.026**
	(0.003)	(0.072)	(0.051)	(0.003)	(0.031)	(0.011)
Observations	80,357	80,357	80,357	80,357	19,671	80,357

Note: This table shows the estimation results for the heterogeneous effects of earthquakes on fertility preferences and son preferences between *earthquake-induced migrant* and *left-behind* women. The dependent variables are regressed as indicator variables for being pregnant, number of pregnancies, number of accumulative children, yearly variation in the number of children, preferences to have daughters in the future and higher son preferences. I regress my outcomes on the variable earthquake and the interaction between earthquake exposure and an indicator variable of migrating after an earthquake, a time-varying measure of covariates, province, age, cohort, and urban at origin fixed-effects (equation (4)). The characteristics included are an indicator variable for being married and being employed. Standard errors are clustered at the district level. The dataset is a person-age panel format (from 1993 to 2014) for women between 15 and 28 when interviewed for the first time. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

F.2.2 Preferences for Education vs Ownership of Physical Assets. Could shifting preferences towards investment in education and away from material possessions drive my results? (S. Becker, Grosfeld, et al. 2020). I examine attitudes toward education and material possessions in Table A.1.42. In Panel A, I use a question from the IFLS rounds 4 and 5 about respondents' parents' expectations about their children's education in the future. In the first four columns, the outcome variable is an indicator that takes the value of one if the expected education is secondary (Columns (1) and (2)) or primary (Columns (3) and (4)). In the last two columns, the outcome is a continuous variable with the expected level of education. Estimates are statistically non-significant, except for the continuous variable. In Panel B, I explore whether there is an actual increase in preferences for education. The outcome variables are as in Panel A but for the actual level of education. Results show that *earthquake-induced migrant women* are more likely to be better educated after the mobility.

I now test whether there is a downward shift in the actual accumulation of assets. In Panel C, my outcome variables take the value one if women's family own non-material (Columns (1)-(2)), material ((3)-(4)) or financial assets ((5)-(6)). I find non-significant results for the first two. Financial assets (i.e. their own or parents' receivables, saving or stocks) have significant results. A possible interpretation of these findings is that displacement may decrease the actual investment in education. However, I do not see a change in asset consumption.[46]

Table A.1.42: Preferences for Education vs. Ownership of Physical Assets

	(1)	(2)	(3)	(4)	(5)	(6)
	PANEL A: Expected Education in the future by their parents					
VARIABLES	secondary educ.	secondary educ.	primary educ.	primary educ.	level educ.	level educ.
*Earthquake * Migration*	0.022	0.013	0.000	0.000	0.122*	0.088
	(0.024)	(0.023)	(0.000)	(0.000)	(0.070)	(0.063)
Observations	26,736	26,736	26,736	26,736	26,736	26,736
	PANEL B: Actual level of education					
VARIABLES	secondary educ.	secondary educ.	primary educ.	primary educ.	level educ.	level educ.
*Earthquake * Migration*	0.033***	0.025**	0.024*	0.022*	0.098***	0.076***
	(0.013)	(0.012)	(0.013)	(0.012)	(0.031)	(0.028)
Observations	215,424	215,424	215,424	215,424	215,424	215,424
	PANEL C: Actual Ownership of Physical Assets					
VARIABLES	non-material	non-material	material	material	financial	financial
*Earthquake * Migration*	-0.031	-0.028	-0.011	-0.011	0.030	0.012
	(0.019)	(0.019)	(0.015)	(0.014)	(0.029)	(0.028)
Observations	146,928	146,928	146,928	146,928	146,928	146,928
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Urban FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes

Note: This table presents the results for the heterogeneous effects of earthquakes on education outcomes and assets ownership between *earthquake-induced migrant* and *left-behind* women. I regress my outcomes on the earthquake variable and the interaction between earthquake exposure and an indicator variable of migrating after an earthquake, a time-varying measure of covariates, province, age, cohort, and urban at origin fixed-effects (equation (4)). The characteristics included are an indicator variable for religion and mother education the previous year of an earthquake. Panel A shows the results on future expectations of children's education by their parents. Panel B reports the estimates on actual education. Panel C presents the physical asset ownership results (non-material, material and financial assets). Standard errors are clustered at the district level. Robust standard errors are in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

F.2.3 Economic Development at Destination Locations. Could my results be driven simply by a move to a place with a more developed education infrastruc-

[46] When my outcome variable is continuous, asset intensity, the results do not change.

ture? To test this potential channel, I employ night light intensity data to measure destination development. I include an interaction term. I do not find a tangible differential effect on the level of education by development at the destination (Table A.1.43).

Table A.1.43: Economic Development at destinations

	Below age 23			Below age 18		
VARIABLES	education	education	education	education	education	education
	(1)	(2)	(3)	(4)	(5)	(6)
*Earthquake * Migration*	0.201***	0.054	0.051	0.082*	-0.004	0.002
	(0.051)	(0.044)	(0.042)	(0.043)	(0.035)	(0.034)
*Earthquake * Migration* * Night light	0.000	0.002	0.001	0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	213,420	213,420	213,420	125,976	125,976	125,976
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Urban FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth Year FE	No	Yes	Yes	No	Yes	Yes
Age FE	No	Yes	Yes	No	Yes	Yes
Controls	No	No	Yes	No	No	Yes

Note: This table presents the results for the heterogeneous effects of earthquakes on education outcomes between *earthquake-induced migrant* and *left-behind* women by economic development at the destination. I use the actual education level as an outcome. I proxy development using night light intensity. I regress education level on the earthquake variable and the interaction between earthquake exposure, an indicator variable of migrating after an earthquake, and an interaction with night light intensity. I include a time-varying measure of covariates, year-island, age, cohort, and urban at origin fixed-effects (equation (4)). The characteristics included are an indicator variable for religion and mother education the previous year of an earthquake. Standard errors are clustered at the district level. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

F.3 Mechanisms: Effects of Earthquake

In this section, I discuss three potential mechanisms of the effects of earthquakes on the annual marriage hazard for the entire population affected. First, bride price is not a determinant factor in marriage decisions. Second, population outflow after an earthquake changes the demographic composition of marriage markets. Third, the destruction of schools may anticipate the marriage of young women.

F.3.1 Bride Price. Bride price means a consumption smoothing channel for households (Corno, Hildebrandt, and Voena 2020). Notably, households hit by a destructive natural disaster may alleviate their financial constraint by acquiring a transfer at the moment of their daughter's marriage. I test this hypothesis by restricting my sample to bride price women. Nonetheless, Panel A of Table A.1.44 shows that a marriage transfer does not affect the results.

F.3.2 Matrilocality. The aggregate labour return of the woman's household increases when newly formed couples live with the bride's family. I check if women in areas affected by earthquakes respond to this economic incentive. Panel B of Table A.1.44 shows that the effects do not hold when restricting the sample to matrilocal women. These findings support the idea that left-behind women do not marry earlier to benefit from consumption smoothing mechanisms because labour markets in earthquake-affected areas are negatively affected in the aftermath of a disaster (Kirchberger 2017; Gignoux and Menéndez 2016).

Table A.1.44: Earthquake effects for bride price and matrilocal women

	Below age 23			Below age 18		
VARIABLES	getting married	get. married	get. married	get. married	get. married	get. married
	(1)	(2)	(3)	(4)	(5)	(6)
			PANEL A: Bride Price Women			
Earthquake	0.004	0.001	0.002	0.001	-0.001	-0.001
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Observations	182,688	182,688	182,688	109,284	109,284	109,284
			PANEL A: Matrilocal Women			
Earthquake	0.055	0.048	0.050	0.066	0.066	0.067
	(0.033)	(0.032)	(0.031)	(0.052)	(0.051)	(0.051)
Observations	28,152	28,152	28,152	16,872	16,872	16,872
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Urban FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth Year FE	No	Yes	Yes	No	Yes	Yes
Age FE	No	Yes	Yes	No	Yes	Yes
Controls	No	No	Yes	No	No	Yes

Note: This Table presents the earthquake results for women engaged in different marriage norms on the dependent variable: annual marriage hazard. The dependent variable is a binary variable for marriage, coded to one if the woman married at the age corresponding to the observation. Earthquakes are defined as earthquakes with an intensity of at least VII in some of its locations affected (Gignoux and Menéndez 2016). Observations are at the level of a person's age at the month level (from 12 to 22 or the age of first marriage). I restrict the sample to women traditionally engaged in the practice of bride price in Panel A and to women traditionally engaged in matrilocal custom in Panel B. The baseline specification is presented in Equation 1. Column (1) presents the results without age, birth year fixed effects, and covariates. Column (2) includes age and birth year fixed effects. Column (3) controls for baseline characteristics (religion and mother education for the year before the earthquake). Columns (4), (5) and (6) perform the same analysis that Columns (1), (2) and (3) but for a sub-sample of ages from 12 to 17 (or age of first marriage). Standard errors are clustered at the district level. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

F.3.3 Social Integration. Local networks can be a key mechanism for coping with the effects of earthquakes. Table A.1.44 shows that local community participation does not affect the results (Columns 3 and 6). However, an important feature of the results is that earthquakes decrease the annual marriage hazard when a household member participates in an *arisan*. An *arisan* is a social club primarily populated by women. Members have similar backgrounds or interests. It represents an alternative to bank loans and other forms of credit. On top of that, the number of *arisan* in which the household participates does not change the results. These findings suggest that the higher financial women's capacity within a household seems to drive the results. In other words, the higher bargaining power of women in deciding their daughter's future changes the direction of the effects.

F.4 Welfare analysis

F.4.1 Early fertility. To estimate the probability of having her first child of the woman i living in district d affected by an earthquake at time t and displaced after it, born in cohort k and having her first child at age a, I use the following baseline specification:

$$Y_{i,d,k,a} = \beta_0 + \beta_1 Eq_{s,p,a,m} + \beta_2 Eq_{s,p,a,m} * Disp_{i,s,a} + Eq_{s,p,a,m} * X_i + \alpha_p + \gamma_a + \delta_k + \zeta_u + \epsilon_{i,d} \tag{1.15}$$

where , $Eq_{s,p,a,m} * Disp_{i,s,a}$, is one if displaced after being exposed to a destructive earthquake in the location of origin s at age a and, 0 otherwise. I further control

Table A.1.45: Earthquake's effects, integration with local population

	Below age 23			Below age 18		
VARIABLES	getting married	get. married	get. married	get. married	get. married	get. married
	(1)	(2)	(3)	(4)	(5)	(6)
Earthquake	0.004	0.005	0.006*	0.012***	0.010**	0.010***
	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
*Earthquake * Arisan*	0.006			-0.010**		
	(0.005)			(0.004)		
*Earthquake * N^o arisan*		0.001		-0.001		
		(0.001)			(0.001)	
*Earthquake * N^o com. act.*			0.000			-0.001
			(0.001)			(0.001)
Observations	585,816	585,816	585,816	350,232	350,232	350,232
Dep. var. mean	0.036	0.036	0.036	0.018	0.018	0.018
Number of provinces	15	15	15	15	15	15
Number of years	22	22	22	22	22	22
Number of districts	255	255	255	255	255	255
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Urban FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth Year FE	No	Yes	Yes	No	Yes	Yes
Age FE	No	Yes	Yes	No	Yes	Yes
Controls	No	No	Yes	No	No	Yes

Note: This Table tests the hypothesis that local networks are a determinant of marriage age. This Table presents the estimates from Equation 1 with interaction to a variable on involvement in local communities: *arisan*, *number of arisan*, and *number of community organizations* women's families participate in. An arisan is a social club that provides its members with alternative bank loans and other forms of credit. The dependent variable is a binary variable for marriage, coded to one if the woman married at the age corresponding to the observation. Observations are at the level of a person's age at the month level (from 12 to 22 or the age of first marriage). Column (1) presents the results with an interaction to a variable being one if a household member participates in an *arisan*. Column (2) includes interaction with the number of *arisan* the household is a member of. Column (3) shows the results with interaction to a variable being one if a household member participates in activities within their community. Columns (4), (5) and (6) perform the same analysis that Columns (1), (2) and (3) but for a sub-sample of ages from 12 to 17 (or age of first marriage). Standard errors are clustered at the district level. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

for a time-varying measure of covariates, $X_{i,a-1}$, in which the woman born in year k is age a, province fixed effects α_p, age, γ_a, year-of- birth fixed effects, δ_k, and urban fixed-effects, ζ_u. Standard errors will be clustered at the district level.

Table A.1.46: Effect on timing of first fertility, *earthquake-induced migrants* vs *left-behind women*

	Below age 23			Below age 18		
VARIABLES	pregnant	pregnant	pregnant	pregnant	pregnant	pregnant
	(1)	(2)	(3)	(4)	(5)	(6)
*Earthquake * Migration*	0.005*	0.003	0.004	0.002	-0.001	-0.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Observations	237,264	237,264	237,264	135,276	135,276	135,276
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Urban FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth Year FE	No	Yes	Yes	No	Yes	Yes
Age FE	No	Yes	Yes	No	Yes	Yes
Controls	No	No	Yes	No	No	Yes

Note: This table displays the estimation results for the effect of earthquakes on the timing of the first child. I regress an indicator variable that takes the value 1 when a woman has her first child on the earthquake variable. I include an interaction between indicator variables for years of exposure and an indicator variable of migrating after an earthquake, a time-varying measure of covariates, province-fixed effects, age-fixed effects, year-of-birth fixed effects, and urban fixed-effects (equation (4)). The characteristics included are an indicator variable for being Muslim and level of education. Standard errors are clustered at the district level. The dataset is in a person-age panel format. Treatment is defined at the year level. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.1.46 shows no heterogeneous effects on the timing of first fertility between *earthquake-induced migrant women* versus *left-behind women.* However, women's marriage anticipates the annual fertility hazard in 43 pp among *earthquake-induced migrant women* (Column (3) of Table A.1.47). It also has effects on fertility before 18 (Column (6) of Table A.1.47).

Table A.1.47: Effect of early marriage on timing of first fertility, *induced-migrant* women

	Below age 23			Below age 18		
VARIABLES	pregnant	pregnant	pregnant	pregnant	pregnant	pregnant
	(1)	(2)	(3)	(4)	(5)	(6)
Marriage	0.420***	0.429***	0.430***	0.306***	0.301***	0.340***
	(0.022)	(0.022)	(0.024)	(0.042)	(0.043)	(0.049)
Observations	98,844	98,844	89,136	53,928	53,928	46,764
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Urban FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth Year FE	No	Yes	Yes	No	Yes	Yes
Age FE	No	Yes	Yes	No	Yes	Yes
Controls	No	No	Yes	No	No	Yes

Note: This table displays the estimation results for the effect of forced displacement on the timing of first child. I regress an indicator variable that takes value 1 when a woman has her first child on the interaction between indicator variables for years of exposure and an indicator variable of migrating after an earthquake, a time-varying measure of covariates, year-island fixed effects, age fixed effects, year-of-birth fixed effects, and urban fixed-effects (equation (3)). The characteristics included are an indicator variable for being Muslim and level of education. Standard errors are clustered at district level. The dataset is a person-year panel format. Treatment is defined at year level. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

F.4.2 Labour integration. An additional consequence of early marriage is the women's exit from the labour market. I evaluate if early marriage affects labour integration for women.

The duration of interest for this analysis is the time between 12 and 22 (or the age entering the labour market for the first time). Using this panel data and sample, I estimate the probability of being employed by woman i living in district d married at time t born in cohort k and entering the labour market for the first time at age a. I restrict the analysis to *earthquake-induced migrant women*. I estimate the following specification:

$$Y_{i,s,k,a} = \beta_0 + \beta_1 Married_{s,a} + X_i + \alpha_p + \gamma_a + \delta_k + \zeta_u + \epsilon_{i,d} \tag{1.16}$$

where $Y_{i,s,k,a}$ is a binary variable coded as 1 at the age the woman is employed and zero otherwise. The married variable, $Married_{s,a}$, switches to one from the occurrence of the first marriage at age a, 0 otherwise. I control for province α_p, age, γ_a, year-of- birth, δ_k, and urban, ζ_u, fixed-effects. I further control for a measure of individual-level covariates, X_i (religion and education). Standard errors are clustered at the district level. Table A.1.48 shows how early marriage decrease the labour integration by 7% for *earthquake-induced migrant women.*

F.4.3 Matching decisions. To study the characteristics of couples that form during displacement, I examine the following specifications for woman i living in district d affected by an earthquake at time t and displaced after it, born in cohort k and married at age τ, I use the following baseline specification:

Table A.1.48: Effect of early marriage on labour integration, *induced-migrant* women

VARIABLES	Employed	Employed	Employed
	(1)	(2)	(3)
Marriage	0.038	-0.076**	-0.071**
	(0.034)	(0.036)	(0.035)
Observations	55,920	55,920	50,832
Province FE	Yes	Yes	Yes
Urban FE	Yes	Yes	Yes
Birth Year FE	No	Yes	Yes
Age FE	No	Yes	Yes
Controls	No	No	Yes

Note: This Table shows the estimates of marriage on employment. I regress an indicator variable that takes the value 1 when a woman is married on a dummy variable for being employed, a time-varying measure of covariates, province fixed effects, age fixed effects, year-of-birth fixed effects, and urban fixed effects. The characteristics included are an indicator variable for being Muslim and level of education. Standard errors are clustered at the district level. The dataset is in a person-age panel format. Treatment is defined at the year level. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

$$Y_{i,d,k,\tau} = \beta_0 + \beta_1 Eq_{s,p,m} + \beta_2 Eq_{s,p,m} * Disp_{i,s} + Eq_{s,p,m} * X_i + \alpha_p + \delta_k + \xi\tau + \zeta u + \epsilon_{i,d} \quad (1.17)$$

In this specification, $Eq_{s,p,m} * Disp_{i,s}$, is one if displaced after being exposed to a destructive earthquake between ages 12 and 23 in the location of origin s at time t and, 0 otherwise. I control for a time-varying measure of covariates, X_i, in which the woman born in year k, province fixed effects α_p, year-of-birth fixed effects, δ_k, year of first marriage, $\xi\tau$, and urban fixed-effects, ζ_u. Standard errors will be clustered at the district level. It is important to notice that we cannot assign any causal interpretation to these estimates, as they are the result of both selection forces (i.e. the characteristics of individuals who chose to marry during a displacement may differ from those who did not) and causal forces (i.e. the fact that a couple married during displacement may lead to different long-term outcomes).

Table A.1.49 shows that *earthquake-induced migrant women* are more likely to have lower education than their spouse, to be in a polygynous marriage, and more likely to marry a *earthquake-induced migrant groom.*

F.4.4 Household consumption capacity. Are household decisions efficient? Do *earthquake-induced migrant women* 's household end up better off after their daughter's marriage? I study how household income and expenditures change after their daughter's marriage to answer these questions. I use data on labour and non-labour income and food and non-food expenditure from the IFLS. I estimate the following specification:

$$Y_{i,s,k,a} = \beta_0 + \beta_1 Married_{s,a} + X_i + \alpha_i + \gamma_a + \delta_k + \zeta_u + \epsilon_{i,d} \quad (1.18)$$

Table A.1.49: Marriage characteristics at the time of marriage, *induced-migrant* women

VARIABLES	edu gap	age gap	polygyny	displaced husb
	(1)	(2)	(3)	(4)
*Earthquake * Migration*	0.143***	0.082	-0.036*	0.131*
	(0.041)	(0.296)	(0.018)	(0.071)
Observations	4,596	7,452	3,540	3,108

Note: This table presents earthquake effects on a set of characteristics on women's marriage: the education gap between spouses (Column (1)), the age gap between spouses (Column (2)), polygyny (Column (3)), and having a spouse who migrated induced by an earthquake. I regress the variable of interest on the earthquake variable and the interaction between the earthquake indicator and an indicator variable of migrating after an earthquake, a time-varying measure of covariates, province, year-of-birth, age at first marriage, and urban fixed effects (equation (1.17)). Standard errors are clustered at the district level. The dataset is a person-age panel format. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

where $Y_{i,s,k,a}$ is a continuous household income and expenditure variable. The married variable, $Married_{s,a}$, switches to one from the occurrence of the first marriage at year t, 0 otherwise. I control for province α_p, age, γ_a, year-of- birth, δ_k, and urban, ζ_u, fixed-effects. I further control for a measure of individual-level covariates, X_i (religion and father's education). Standard errors are clustered at the district level. Table A.1.50 presents the results. Their daughter's marriage seems not to affect labour income and expenditures. But, non-labour income decreases.

Table A.1.50: Effect of early marriage on household's welfare, *induced-migrant* women

	PANEL A. Household income					
VARIABLES	labour	labour	labour	non-labour	non-labour	non-labour
	(1)	(2)	(3)	(4)	(5)	(6)
Marriage	2406613.129***	398,057.048	370,812.266	-81,713.697***	-68,333.030*	-64,193.105*
	(603,288.241)	(739,712.415)	(715,834.606)	(28,981.203)	(37,638.718)	(38,663.973)
Observations	39,204	39,204	39,204	78,228	78,228	78,228
	PANEL B. Household expenditure					
VARIABLES	food	food	food	non-food	non-food	non-food
	(1)	(2)	(3)	(4)	(5)	(6)
Marriage	23,771.999***	1,376.088	-289.335	-81,364.646	-1295981.490	-1434933.824
	(7,834.210)	(9,947.414)	(10,003.991)	(701,235.517)	(1034763.300)	(1078381.069)
Observations	78,276	78,276	78,276	78,276	78,276	78,276
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Urban FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth Year FE	No	Yes	Yes	No	Yes	Yes
Age FE	No	Yes	Yes	No	Yes	Yes
Controls	No	No	Yes	No	No	Yes

Note: This table shows the equation estimates (1.18). The dataset is in a person-year panel format. Treatment is defined at the year level. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

F.5 Forced displacement: An Income shock

In this section, I evaluate if the effect of earthquakes on income is different between *earthquake-induced migrant* (forcibly displaced) and *left-behind* population. I proxy income using labour market outcomes. As in the main analysis, I exploit random geographic and time variation in the occurrence of earthquakes to implement a difference-in-differences strategy in a duration model.

The duration of interest for this analysis is the time between 15 and 45 (or the age entering the labour market for the first time), the standard definition of the active population. I convert my data into a person-year-month panel format. To later on, merge these individual data with earthquake data at the year-month level and covariates at the year level.

Using this panel data and sample, I estimate the probability of being employed of individual i living in district d affected by an earthquake at time t born in cohort k and entering the labour market for the first time at age a. I restrict the analysis to the population exposed to earthquakes and compare displaced to stayers women. I estimate the following specification:

$$Y_{i,s,k,a} = \beta_0 + \beta_1 Eq_{s,p,a,m} + \beta_2 Eq_{s,p,a,m} * Disp_{i,s,a} + Eq_{s,p,a,m} * X_i + \alpha_p + \gamma_a + \delta_k + \zeta_{u_o} + \epsilon_{i,d} \tag{1.19}$$

where $Y_{i,s,k,a}$ is a binary variable coded as 1 at the age the individual is employed and zero otherwise. The exposure to an earthquake, $Eq_{s,p,a,m}$, switches to one from the occurrence of an earthquake in the sub-district of resident s at the age a, 0 otherwise. $Disp_{i,s,a}$ switch to 1 if displaced after the shock $Eq_{s,p,a,m}$. I control for year-island fixed effects α_p, age fixed effects, γ_a, year-of- birth fixed effects, δ_k, and urban at origin, ζ_u, fixed-effects. I further control for individual-level covariates measured a year before an earthquake strikes, X_i (mother education and religion). Standard errors are clustered at the district level.

Table A.1.51 shows how displacement decreases the annual hazard of being employed. In column 3, I report the estimated coefficients for equation 18. It shows that individuals who experience an earthquake between the ages 15 and 45 are 0.5 percentage points (pp) less likely to get employed in the same year. The effect is statistically significant at the 1% level. The average annual marriage hazard for this age group is equal to 0.919. Hence, the effect corresponds to an approximately 5% decrease in the annual employment hazard in response to an earthquake.

Column 6 shows the results for a sub-sample of women. Women who experience an earthquake between the ages 15 and 45 are 0.4 percentage points (pp) less likely to get employed in the same year. The effect is statistically significant at the 5% level. The average annual marriage hazard for this age group is equal to 0.813. Hence, the effect corresponds to an approximately 5% decrease in the annual employment hazard.

Table A.1.51: Earthquake effects on labour, *induced migrants* vs *left-behind* population

	All			Women		
VARIABLES	working	working	working	working	working	working
	(1)	(2)	(3)	(4)	(5)	(6)
*Earthquake * Migration*	-0.057***	-0.039***	-0.049***	-0.081***	-0.039**	-0.039**
	(0.015)	(0.013)	(0.013)	(0.017)	(0.016)	(0.016)
Observations	514,972	514,972	514,072	242,768	242,768	242,429
Mean	0.919	0.919	0.919	0.813	0.813	0.813
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Urban FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth Year FE	No	Yes	Yes	No	Yes	Yes
Age FE	No	Yes	Yes	No	Yes	Yes
Controls	No	No	Yes	No	No	Yes

Note: This Table presents the effects of earthquakes on a labour outcome between textitearthquake-induced migrants vs *left-behind* individuals. The sample includes an active population (from 15 to 45). I regress an indicator variable that takes the value 1 when working (0, otherwise) on the earthquake variable. I also add an interaction between an earthquake variable, $Eq_{s,a}$, and an indicator variable of migrating after an earthquake, $Disp_{s,a}$. I include a time-varying measure of covariates, province-fixed effects, age-fixed effects, year-of-birth fixed effects, and urban fixed effects. The characteristic included is having primary education. Standard errors are clustered at the district level. Columns (4)-(6) run the same analysis for a subsample of women. The dataset is in a person-age panel format. Treatment is defined at age level. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

G Appendix for Theoretical framework

G.1 Proof of propositions

Proof of proposition 1. A household i wants its daughter to get married by the end of the second period if and only if:

$$\ln(y_2(1-d_2)+\epsilon_2^w+w^f+p_2+w_2^m)+\eta_2+\xi^f > \ln(y_2(1-d_2)+\epsilon_2^w+w^f)$$

$$\iff p_2 \geq \frac{(1-exp(\eta_2+\xi^f))}{exp(\eta_2+\xi^f)}\ (y_2(1-d_2)\ +\epsilon_2^w\ +w^f) - w_2^m = \underline{p_2}$$

For household j with a son, we follow similar algebra:

$$\ln(y_2(1-d_2)+\epsilon_2^m-w_2^m-p_2)+\eta_2 > \ln(y_2(1-d_2)+\epsilon_2^m+w_2^m)$$

$$\iff p_2 \leq \frac{(exp(\eta_2)-1)}{exp(\eta_2)}\ (y_2(1-d_2)\ +\epsilon_2^m) - \frac{(exp(\eta_2)+1)}{exp(\eta_2)}\ w_2^m = \overline{p_2}$$

Proof of proposition 2. The derivative of the equilibrium price in the supply with respect to income is equal to

$$\frac{\partial p_1}{\partial y_1} = \frac{(1-d)[(1-\eta_1^f)(\eta_1^m-1)+(1-\eta_1^m)(\eta_1^f-1)]}{2-\eta_1^m-\eta_1^f}$$

The derivative is positive when $\eta_1^m + \eta_1^f < 0$, and, negative when $\eta_1^m + \eta_1^f > 0$.

Proof of proposition 3. The derivative of the equilibrium quantity in the supply with respect to income is equal to

$$\frac{\partial Q(y_1)}{\partial y_1} = \frac{(1-d)[\phi_1(\eta_1^f-1)-\phi_2^2[\phi_2(\eta_1^m-1)+(1-\eta_1^m)(\eta_1^f-1)]]}{(\phi_1\phi_2)^2}$$

The sign of the derivative is ambiguous. It is positive or negative depending if the value of $\eta_1^m + \eta_1^f$ is $>$ *or* $<$ than 2 and if η_1^f is $>$ *or* $<$ than 1.

Proof of proposition 4. The derivative of the equilibrium quantity in the supply with respect to the groom's labor contribution, is equal to

$$\frac{\partial Q(w_1^m)}{\partial w_1^m} = \frac{-[\phi_1+2\phi_2^2]}{(\phi_1\phi_2)}$$

The sign of the derivative is ambiguous. It is positive or negative depending if the value of $\eta_1^m + \eta_1^f$ is $>$ *or* $<$ than 2 and if η_1^f is $>$ *or* $<$ than 1.

G.2 Displacement and equilibrium in an aggregate market

Displacement as an economic shock. Displacement is an unexpected income shock for forcibly displaced households. I assume that the shock turns into a reduction in household income by fraction d. This shock affects displaced households with daughters or sons. But, it does not affect the *native* households in the marriage market at the new destination. Thus, it implies that the supply of child brides increases in displacement, and the demand for child brides decreases among displaced households and is unchanged among *native* households.

The equilibrium bride price will change by $\frac{-y_1 d[\phi_2(\eta_1^m-1)+(1-\eta_1^m)(\eta_1^f-1)]}{\phi_1}$. The effect of displacement on the price of child marriages is ambiguous and increases only if $\eta_1^f + \eta_1^m < 2$ and $\eta_1^m \neq 0$ or $\eta_1^m = 0$ and η_1^f does not range from (1,2). The equilibrium number of child marriages will increase, as a result of displacement, by $\frac{y_1 d[\phi_1(\eta_1^f-1)-\phi_2^2[\phi_2(\eta_1^m-1)+(1-\eta_1^m)(\eta_1^f-1)]]}{\phi_1\phi_2}$.[47] [48].

Responsiveness to bride price. The net change in the equilibrium number of child marriages will depend on the relative responsiveness of the supply and the demand for child brides when the equilibrium bride price decreases. Figure A.1.12 illustrates two possible scenarios that might result in an equilibrium. If the supply curve (S) is steeper than the demand curve (D), the number of child marriages will

[47] Always that $\phi_1(\eta_1^f - 1) > \phi_2^2[\phi_2(\eta_1^m - 1) + (1 - \eta_1^m$

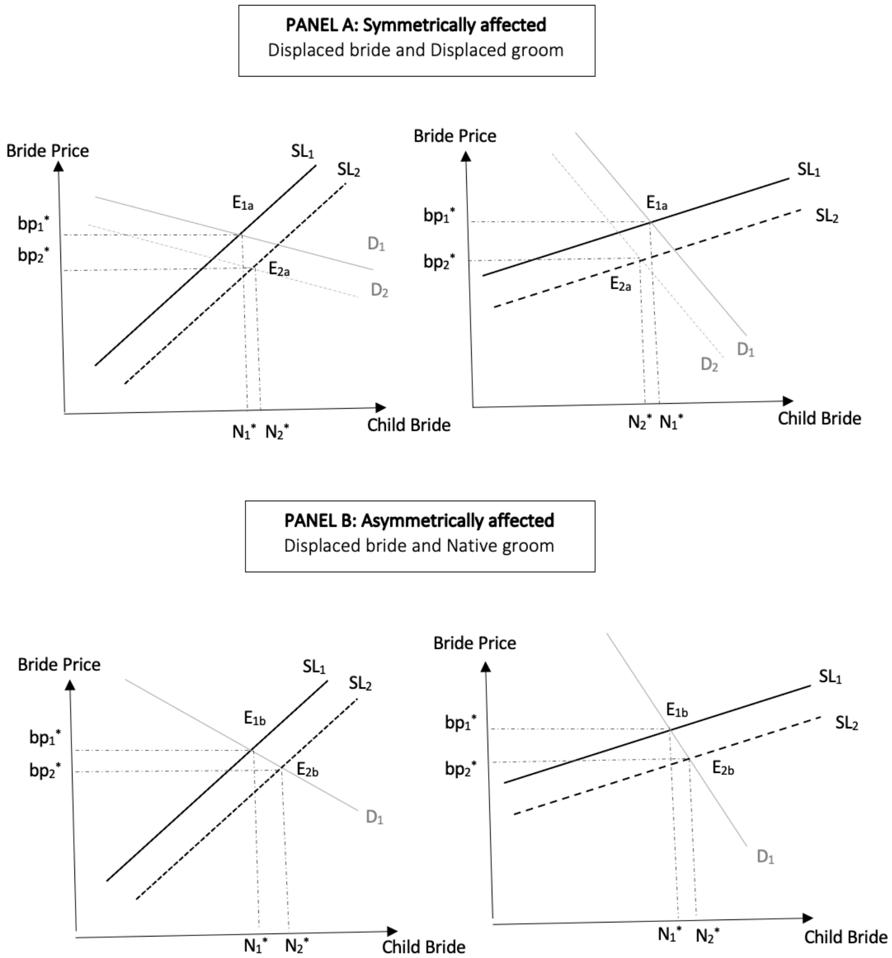
[48] Note that the slope of the supply of child brides is $\frac{1}{1-\eta_1^f}$ and the slope of the demand for child bride is $-[\frac{1}{1-\eta_1^m}]$

increase from (N_1^*) to (N_2^*) starting from the initial equilibrium at (E_{1a}) or (E_{1b}) (left-side graphs of panel A and B of Figure A.1.12). If the supply curve (S) is flatter than the demand curve (D), the number of child marriages will decrease from (N_1^*) to (N_2^*) at the new equilibrium (E_{2a}) between displaced brides and grooms (right-side graph of panel A of Figure A.1.12). Nevertheless, at the new equilibrium (E_{2b}), where the supply of displaced brides meets the demand by *natives*, the number of child marriages will increase from (N_1^*) to (N_2^*) but in much smaller magnitude (right-side graph of panel B of Figure A.1.12).

The financial capacity of displaced households is substantially affected by their displacement. Therefore, the supply of child brides is more likely to be less price elastic than its demand in a matrilocal setting since the bride's household may strongly rely on their son-in-law's labour return (and their daughter's home support).

However, the average effects on child marriage in the marriage market in the new destination will depend on the compositional effects of each demand. Since the direction of the effect of displacement on child marriages is theoretically ambiguous, it is a matter of empirical inquiry.

Figure A.1.12: Two possible scenarios for an equilibrium



Note: This Figure shows two possible equilibriums in the aggregate marriage market. Panel A presents the equilibrium for matching between the displaced bride and displaced groom. Panel B presents the equilibrium for couples of the displaced bride and *native* groom.

Chapter 2

Reservoir-induced displacement and social participation: Evidence from the Spanish Dictatorship

Revised & Resubmitted at the European Economic Review

2.1 Introduction

By 2021, 89.3 million people worldwide had been forced to flee from their homes (UNHCR 2022). Among them, 60% were internally displaced persons.[1] Despite the growing research examining the impacts of forcibly displaced population inflows in hosting communities, very little attention has received the social participation consequences. Do large inflows of internally displaced persons cause long-lasting social participation changes in hosting locations? Examining these effects is crucial given the direct consequences of social participation on the economic (Guiso and et.al 2004); education (Coleman 1988) and health implications (Carpiano 2008) for the populations living inside hosting communities.

To answer this question, we need a plausibly exogenous variation to capture displaced population flows, which is uncommon in many settings. In line with the striking estimates of 15 million people displaced every year by infrastructure projects (e.g. transport, mining, and water projects) (IDMC 2017), this paper overcomes this empirical challenge by measuring forced displacement as internal displacement associated with water infrastructure projects. Reservoirs are public

[1] Internally displaced persons (IDPs) are 'persons or groups of persons who have been forced or obliged to flee or to leave their homes or places of habitual residence, in particular as a result of or to avoid the effects of armed conflict, situations of generalised violence, violations of human rights or natural or human-made disasters, and who have not crossed an internationally recognised state border.' (UN 1998)

goods investments.[2] Nevertheless, even if a large population benefits from these reservoirs' water and energy services, their development comes at a cost. Globally, the construction of reservoirs has led to the displacement of 40-80 million people (WCLD 2000).

I examine how exposure to internally displaced population inflows that happened in the past affected social participation in host municipalities during the next 50 years. I capture exogenous variation in internally displaced persons flows by exploiting a unique historical setting -the construction of reservoirs during the Spanish dictatorship (1936 to 1975) in the Ebro's river catchment area- and a newly-collected historical panel data on internal displacement associated with reservoirs and social participation. I measure social participation with voter turnout (general and municipal elections), total cooperatives, and non-profit associations from 1977 to 2018.

Empirically, I look at municipalities that received internally displaced persons. I call them *host* municipalities. They are the municipalities adjacent to the destroyed municipalities by reservoir construction. To estimate the effects of a reservoir-induced displaced population inflow, municipalities adjacent to municipalities with a reservoir, whose construction did not destroy any villages, function as my control group. Notably, the presence of this counterfactual allows for the isolation of the impacts from a reservoir (Duflo and Pande 2007). Nonetheless, potential omitted variables related to the location of reservoirs, the destruction of a village, and the destination selection could encapsulate endogeneity problems. To address these challenges, I implement an instrumental variable strategy by combining three sources of variation. First, to overcome the non-randomness of the reservoir's location, I exploit the margin of whether a reservoir was planned or not before the dictatorship. Second, I use the reservoir size that strongly affects the probability of a village being destroyed. Third, I profit from marginal variation in the distance to each adjacent municipality to upfront the potential self-selection into a destination.

The main results show that the municipalities that *Host* internally displaced persons have a long-term effect on the decrease in their social participation. In particular, *hosting* forcibly displaced population decreases the voter turnout in the general and municipal elections by 15 and 13 percentage points with respect to the municipalities adjacent to municipalities with a reservoir that did not destroy a village (the control group). The effects are significant at the 1% and 10% level and are sizable in economic magnitude, corresponding to 20% and 17% decrease in the outcomes mean in 1977-2019. There are also statistically significant effects on the

[2] In many cases, the benefits of water infrastructure projects are both 'non-rival' and 'non-excludable'. For instance, when a reservoir benefits the population adjacent to the river from the reduced risk of flooding. Moreover, increasing the number of people benefiting from the reservoir does not affect resource availability.

number of non-profit associations established. *Host* municipalities are 0.8 less likely to have an additional association per capita from 1977 to 2018 than their control group. The average number of associations in the municipalities in my sample is three. Therefore, the magnitude of the effect is quite relevant. The effects on the total number of cooperatives are statistically non-significant.

To study whether differential patterns in social participation exist when the intensity of the treatment changes, I examine how the relative weight of the internally displaced population affects the results. I show that the number of the forcibly displaced population relative to the native mitigates the impacts in general elections participation, municipal elections turnout and the number of associations by a 20, 15 and 15%, respectively. The higher the forcibly displaced population who arrive in a municipality, the smaller the reduction in social participation. I also explore when the effects on social participation started. The negative impacts on the number of associations started from the arrival of the forcibly displaced population in the *Host* locations. I document that the impacts persisted after the dictatorship for over 50 years.

The decline in institutional and general trust is the underlying mechanism behind the main results. First, effective policy making can lead to an increase in trust in the government (Fair et al. 2017; León-Ciliotta, Zejcirovic, and Fernandez 2023), and as a consequence, affect political participation (Putnam 1995). I find a 42% increase in mistrust of democratic institutions and three times higher vote shares for Falangist parties in the general elections in *host* municipalities. In turn, a higher share of the internally displaced population compared to the native population in the treatment year mitigates both of the estimates. These findings suggest that voters accounted for the effectiveness of the government responsible for the construction with two opposing behaviours: whereas the natives rewarded, the internally displaced population punished the government of Franco. Their gratitude and displeasure have persisted decades later in their actions. Second, inter-group conflict leads to changes in general trust. Whereas the within-group bond strengthens and individuals become more cooperative within that group, the inter-group trust decreases (Bauer et al. 2016). I show that *host* communities raise their intra-group trust by 32% and decreases in 23% the inter-group trust between native and forcibly displaced population compared to the control municipalities. These results indicate that the conflict between the citizens positively and negatively affected by the same reservoir fostered distrust between the two groups within the *Host* communities, impacting their social cohesion and persistence over time.

In order to support the validity of my results, I generate empirical evidence that indicates no statistical relationship between the instrument (or any of their variables) and social participation in the parts of the Ebro region where the destruction of villages did not occur. The exercise supports the validity of the exclusion restriction. I also test whether my results are sensitive to my choice of measure for *Host* municipality by expanding the definition to the surrounding municipalities to the *Host* municipalities. Additionally, I rule out the confounding effect of the violence perpetrated during the dictatorship and local development.

Despite the historical setting, the research question remains relevant today. Internal displacement resulting from water infrastructures is a reality that affects many communities worldwide. Considering the recent displacements induced by the Son La Dam (Vietnam), Gilgel Gibe III Dam (Ethiopia-Kenya), and Ituango Dam (Colombia), it is evident that the toll of water infrastructure projects is still on the rise. Additionally, my results highlight the importance of responding to sizeable sudden migration waves by increasing the inter-group cohesion between native and forcibly displaced populations. Neglecting to respond to forcibly displaced population's integration may end up hurting social participation in *Host* communities, with long-lasting consequences over the following decades, and ultimately negatively impacting welfare.

This paper contributes to three strands of the literature. First, it belongs to the literature that studies the long-term impacts of the forcibly displaced population in hosting communities. Most of the existing literature has focused on economic (Alix-Garcia, Walker, et al. 2018; E. Murard and Sakalli 2021; Arbatli and Gokmen 2023) and educational outcomes (Bharadwaja and Mirzab 2019; Morales 2018). I contribute to the literature by constructing a new historical panel database about forced displacement due to reservoir construction that other researchers can use. The data-set enables me to identify the municipalities destroyed and the population affected and identify some of the *receiving* municipalities. Using this information, I can further contribute to this research agenda by investigating the long-lasting consequences of a short-distance migration on social participation. Literature on this type of migration is scarce, and we know very little about the long-term effects of short-distance migration on social participation.

Second, this study contributes to the literature on the long-term determinants of social participation (Cagé and Rueda 2016). It exists some research on the effects of migration inflows on social participation. For instance, Levy 2018 shows how mass migration reduces community organisational life and community volunteerism in the short-term. However, the singular setting of this paper allows me to contribute to this literature by presenting new evidence on the consequences of the arrival of a large population forcibly relocated on social participation in *Host* communities fifty years after. The closest paper to this study is Abel 2019, which studies the long-lasting effect of forced removals on migrants' level of trust in South Africa.

Finally, this paper adds to the literature studying the consequences of infrastructure projects using a credible counterfactual. Among the earlier attempts are the works of Duflo and Pande 2007 and Asher and et.al 2022. I contribute to this literature by quantifying the negative externality of public goods provision on out-migration.[3]

The rest of the paper is organised as follows. Section 2 provides an overview of the historical background. Section 3 presents the data used. Section 4 explains the empirical strategy. Section 5 and 6 discuss the main results and their robustness, respectively. Section 7 evaluates the mechanisms. Section 8 presents the conclusions.

2.2 Historical Background

Although displacements associated with reservoir buildings persisted during the democracy (1976-today), the number of dislocations after 1975 (the end of the Spanish dictatorship) was very few.[4] Actually, it is not accidental that this paper's setting falls into the Spanish dictatorship. Only under an authoritarian regime is it possible to imagine a setting capturing different dislocations of communities due to multiple reservoir buildings. In this section, I provide some historical accounts to explain what made this singular displacement setting possible.

2.2.1 The Spanish dictatorship

In July 1936, part of the Spanish military stationed a coup against the Republican government (1931-1936). The military's coup led to the beginning of the Spanish Civil War (1936-1939), which drove Spain to the outset of the Spanish dictatorship.[5] From 1936 to 1975, a fascist dictatorship ruled Spain until Franco's death in November 1975. Subsequently, a democracy (1976-today) was established. See Figure A.2.2 for a historical timeline.

2.2.2 Water policy in Spain: Reservoirs' construction

Spain has the second-highest number of dams per km^2 in the world, 0.23, and is the fifth country with the most dams and reservoirs worldwide[6]. These figures are a direct consequence of a period of meticulous and exhaustive projection of

[3] Duflo and Pande 2007 analyses the unequal cost-benefit and regional distribution of dams and reservoir construction.Asher and et.al 2022 estimate the long-run impacts of India's irrigation canals on the increase in agricultural productivity and population density

[4] In Ebro's four reservoirs (Alba, Itoiz, Pajaras, and Rialb) displaced the population after 1975.

[5] In the Civil War, Republicans loyal to the left-leaning Popular Front government of the Second Spanish Republic fought against a revolt by the Nationalists.

[6] With 1,200 reservoirs of at least 15 m in height, following China, USA, India, and Japan (Romero 2013)

water infrastructures during the Second Republic (1931-1936) and decades with the construction of a colossal number of reservoirs during the dictatorship (1936-1975). See Figure A.2.3 for a description of the number of dams and reservoirs over time.

In 1933, the Second Spanish Republic laboriously designed the Water Infrastructures National Plan (*Plan Nacional de Obras Hidraúlicas*), which projected the reservoirs across the different catchment areas in Spain.[7] Three years later, in 1936, the Civil War started, leading to the beginning of the dictatorship. Nonetheless, Franco's government used the 1933 Plan (as I refer to *Plan Nacional de Obras Hidraúlicas* in the paper) to design their Water Policy, becoming the water infrastructure planning tool for the future decades. Indeed, a considerable number of infrastructures projected in the 1933 Plan were built during the dictatorship (Moral 2009).

The use of low-priced labour and prisoners, the funding entirely provided by the government, the lack of environmental laws, and the nonexistence of a mechanism of defence for the population affected facilitated the boost in reservoir construction during the dictatorship. Through the construction of reservoirs, the dictatorship aspired to meet the growing urban water supply, increase the productive capacity of the agricultural sector, as well as to consolidate Franco's reputation in rural Spain after the Civil War (1936-1939) (Romero 2013).

Case study: The Ebro's river catchment area. In this paper, I focus on the Ebro's river catchment area, the most extensive Spanish catchment area (17 % of the national territory).[8] Historically, many reservoirs have been developed in this region due to their geographic characteristics. Ebro is the largest river in Spain. I take advantage of this fact and use granular historical information on reservoirs to identify the villages demolished by these constructions. See the map of Figure 2.1 for further details.

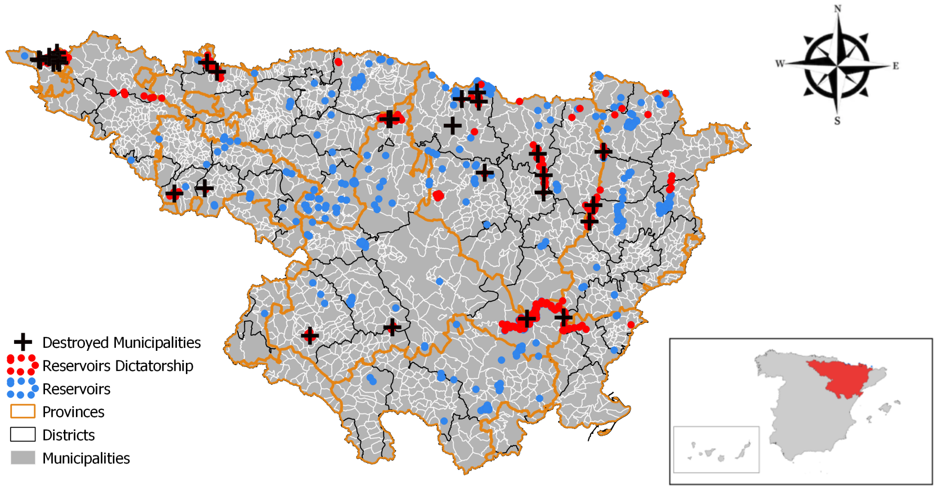
2.2.3 Reservoir-induced displacement during the Spanish dictatorship

Reservoirs' construction during the 20th century had submerged at least 500 villages in Spain, with an estimate of 50,000 internally displaced persons (EiA 2018). Just in the Ebro area, reservoirs displaced at least 21,561 people. Importantly for my

[7] Although reservoir construction started in the first century with the Romans, the critical water infrastructure ideology in Spain began in 1820. Beyond that, there were previous attempts at water infrastructure projects. For instance, the Gasset Plan (1902) had projected 300 water infrastructures (but only 30 were executed before 1933).

[8] The Ebro's catchment area is on the North-East of the Iberian Peninsula, with a total area of 85,534 km2. Its natural borders are: on the North by the Cantabrian Mountains and the Pyrenees; on the South-East by the Iberian Mountain System; and on the East by the Catalan Coastal Range chain.

Figure 2.1: Reservoirs in Ebro's river catchment area (my case study)



Note: This map shows the spatial distribution of reservoirs in the Ebro region, my case study. Red polygons correspond to reservoirs built during the Spanish dictatorship (1936-1975). Black crosses illustrate the destroyed municipalities by reservoirs during the dictatorship. Blue polygons correspond to reservoirs built before 1936 or after 1975. The smaller map on the bottom right side of the figure shows Ebro's catchment area in Spain. Source: Inventory of Dams and reservoir dataset. Spanish Ministry for Ecological Transition

identification, 67% of the reservoirs that generated forced displacement during the dictatorship was already projected by the 1933 Plan.

Notably, during the Spanish dictatorship, the construction of reservoirs disregarded internally displaced persons' rights (as stated by Fundación Cerezales and Desplazados' project during my interviews). Although there was occasional cash compensation for the affected population, it was minimal, and they suffered from delays and fraud (Marcos and Fernández 2019). Figure 2.2 illustrates one example of dislocation by a reservoir in Ebro's region. It shows how the population of Fayón (a village in Zaragoza province) was forced to leave their houses due to the water level increase created by the Ribaroja reservoir in 1967. A new village was built some kilometres from the old Fayón. Nevertheless, there were very few villages which were rebuilt. Most of the affected population moved to close municipalities. For instance, the population of Jánovas (a village in Huesca province destroyed by a reservoir which was never built) moved to municipalities nearby, such as Sabiñánigo (40 kilometres from Jánovas) (Marcos and Fernández 2019).

The affected neighbours received letters giving them the date and time for the eviction and the expropriation of their properties. The time between the notification and the moment they actually moved to a new municipality depended on neighbours' resistance. A local newspaper gathers testimony from some of the neighbours from Fayón. The ex-major of Fayón said, "*When this letter arrived, you had no choice but to comply and have everything prepared and packed.*". An old woman who was 17 when the relocation said, "*It marked us all because it was a total life change. Before,*

Figure 2.2: Population relocation in Fayón (Zaragoza), 1967, by Ribaroja reservoir

Note: This picture illustrates the displacement process due to the increasing water level of the Ribaroja reservoir in 1967 in Fayón. Fayón is a municipality in Zaragoza province. Source: Aragon TV, 2017.

you had some neighbours, and now there are others. The relationship before the eviction was excellent and familiar, but that divided us between those who accepted the 50,000 pesetas in compensation, gave up having a house, and left and those who stayed. The police took us out to the force.".

The reservoirs submerged most of the villages entirely. Figure 2.3 shows the spatial destruction of Artozki (Navarra), a village that was destroyed, in 2003, by constructing the Itoiz reservoir. A limited number of villages were only partially destroyed. Tiermas and Ruesta (Zaragoza) are examples of the latter. They were partially expropriated by Yesa Reservoir in 1948.

In a nutshell, in this paper, I profit from a Republican plan (1933 plan) to design my identification strategy, the Spanish dictatorship (1936-1975) as a treatment period, and the democratic period (1976-2018) as my outcome period. See the timeline in Figure A.2.2 for a visual summary.

2.3 Data

I assemble a novel panel dataset at the municipal level that combines historical data on reservoirs, villages destroyed by reservoirs and internally displaced persons flows with historical and contemporary social participation outcomes.

2.3.1 Dams and reservoir data

I obtain data on reservoirs in Spain from 1903 to 2010 from the inventory of dams and reservoir dataset from the Spanish Ministry for Ecological Transition. It contains

Figure 2.3: Municipalities destruction, Artozki example by Itoiz Reservoir in 2003

Note: These pictures illustrate the destruction of Artozki (in Navarra province) due to a Itoiz reservoir in 2003. Source: Valle de Arce-Artzibar website.

geo-referenced information on reservoir location, finalisation year, reservoir size, reservoir's primary goal, and other features.

The reservoir stock in Ebro's catchment area equals 130 reservoirs. The dictatorship (1936-1975) built 52 (40% of the total). There is a concentration of reservoirs in the Northern province of Huesca (38%), Lleida (17%), and Zaragoza (8%). The blue and red polygons in Figure 2.1 represent the reservoirs in Ebro's region.

Additionally, I have extracted from the 1933 Plan (or *Plan Nacional de Obras Hidraúlicas*) the reservoirs which were planned to build, and those which were rejected or understudy. This information allows me to identify among the reservoirs built in the dictatorship which reservoirs were already planned before Franco. 32% of reservoirs built during the dictatorship were already planned in 1933 Plan.[9]

[9] Reservoirs projected in 1933's Plan but built during the dictatorship: El Grado, Ribaroja, Fayón, Canelles, Yesa, La Tranquera, Las Torcas, Santa Ana, Mediano, Mansilla, Ebro, Puente La Reina, Sariñena, González Lacasa, Vadiello, Oliana, Flix, and Jánovas

2.3.2 Reservoir-induced displacement data

One of the major contributions of this paper is to build a unique geo-referenced dataset on the forced movement of the population created by the construction of reservoirs in Ebro's region from 1936 to 1975. I first identify the municipalities destroyed by each construction to create this dataset. The information comes from qualitative information in a text or excel format from old local organisations, community associations (created after a village extinction by reservoirs' construction) and digitised local newspapers. This dataset covers all 1,695 municipalities in Ebro's river catchment area during the dictatorship. Table A.2.1 describe the data sources in detail.

On top of finding out which villages were demolished by water infrastructures, I estimate the total affected population at the municipal level, the year of displacement, the level of destruction of a village (entirely or partially destroyed), and actions in place (municipality reconstruction or resettlements). I know some of the municipalities of destinations where most of the population moves to. However, I do not have information on the total population that arrived at each municipality (called *host* municipalities; see section 4.1 for more details). Although this is an important limitation. There is anecdotal and historical evidence that shows that most of the affected population by a destroyed municipality settle down in the surrounding municipalities. I give further details on the database construction in the data Appendix where I define each variable, how I measure them, and their limitation. In particular, in Table A.2.1 I describe the destinations of the affected population with their respective data sources.

Ninety-four municipalities were affected by the construction of reservoirs during the dictatorship in Ebro's area, generating approximately 21,561 internally displaced persons.[10] The red polygons in Figure 2.1 show the (17) reservoirs which destroyed municipalities during the dictatorship.[11]

Considering that one of the main constraints to studying the effect of reservoir-induced displacement in Spain is the lack of official statistics and consistent estimates, this paper's data contribution will be useful for future research.[12]

[10] These figures do not include the people displaced by other aspects of the projects nor those displaced in years after reservoir completion.

[11] Búbal, Canelles, Ebro, Escales, Gónzalez Lacasa, El Grado, Lanuza, Mansilla, Mediano, Mequinenza, Ribaroja, Santa Ana, La Tranquera, Ullivarri, Yesa, Jánovas, and Las Torcas

[12] (Romero 2013) estimates 25,000 people displaced in Spain due to reservoirs. (EiA 2018) counts for at least 500 villages devastated in Spain as a result of reservoirs during the 20th century, displaying 50,000 people. Whereas Ebro river's Hydraulic Confederation (which is responsible for reservoirs construction in Ebro's region) estimates only 13,000 displaced by all reservoirs built throughout history in Ebro, the Commission of Population Affected by Large Reservoirs (COAGRET) estimates 12,000 displaced only in the Aragon region (50% of Ebro's region).

2.3.3 Social participation data

I use five different measures of social participation at the municipal level. The first two measures are publicly available voter turnout in the general and municipal elections, published by the Spanish Ministry of Internal Affairs. This is in line with Cagé and Rueda 2016 and Bellows and Miguel 2009. I have collected data for most of the elections conducted in Spain since democracy was reestablished following Franco's death in 1975. I include information on general elections for the national parliament (Congress) in 1977, 1979, 1982, 1986, 1989, 1993, 1996, 2000, 2004, 2008, 2011, 2015, 2016 and 2019. I complement this data with municipal elections for 1979, 1983, 1987, 1991, 1995, 1999, 2003, 2007, 2011, 2015 and 2019. During the general elections held in the Second Republic in 1931, 1933 and 1936, the Spanish electoral circumscriptions were the province after modifying some articles of the 1907 electoral law. Thus, the official electoral results are not available at the municipal level. [13]

My third measure is the number of cooperatives created. Each Cooperative Register in each region (*Comunidad Autónoma*) has provided this data from 1945 to 2018. I restrict cooperative outcomes from 1977 to 2018 to allow comparability across outcomes. I use the information I have on the economic sector of each cooperative to build my fourth social participation outcome, the number of cooperatives established in the agricultural sector. The primary role of reservoirs in Spain has been hydroelectric power production, storage for public water supply and irrigation. Hence, the impacts on cooperatives in the agricultural sector could be different.

I also use the number of non-profit associations created from 1977 to 2018 to measure social participation. This data comes from the Spanish regional registers of associations. These data are also available yearly from 1945 to 1976.

2.3.4 Trust data

I add data on trust in institutions and other persons from the Spanish Sociological Research Center (CIS, Centro de Investigaciones Sociológicas) to shed light on the mechanisms. The data I have are at the municipal level from eleven cross-sectional surveys conducted from 1989 to 2015.

The specific question related to institutional trust asked in the surveys is the following: Which are the reasons for not voting in the general elections?. The answers to this question are: 1) no alternative satisfies me; 2) I am sick of politics and elections; 3) It does not matter whether to vote or not to vote. It is useless; 4) I am confused, I do not know whom to vote for; 5) Neither parties nor any politician

[13] Some papers have gathered information on voting turnout from official registers of voters for big cities such as Barcelona (Amat and et.al 2020). This information is not available for smaller places.

inspires confidence in me; 6) To show my discontent; 7) For health, work, family reasons; 8) I am not registered; 9) Others. First, I construct a dummy variable called institutional mistrust, which is equal to one if selecting any of the options 1, 2, 3, 5, or 6, zero otherwise. Then, I calculate the average share of individuals who reported mistrust in the national government for each municipality.

Concerning the information on general trust, the specific question related to trust in other persons asked in the surveys is the following Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?. Answers range between 10 (most people can be trusted) and 0 (must be very careful). In a similar vein to the institutional mistrust variable, I measure the average general trust for each municipality.

CIS data are at the individual level. The surveys are representative at the province and primary city levels. However, the sample size does not allow me to exploit this feature. Most of the municipalities in my sample are smaller than 100.000 inhabitants. Therefore, the sample is not representative at the municipal level. To overcome this limitation, I have aggregated all the individual answers registered in the eleven surveys within a given municipality. Even conducting this exercise, I could only obtain data for 266 municipalities.

2.3.5 Other municipal level data

I include administrative data on municipalities (N=1,695), provinces (N=18), and regions (or *Comunidad Autónoma*) (N=8) from the Ebro basin's Hydrographic Confederation (ChE). I also add information on each province's districts (or *comarcas*). A *comarca* is a group of municipalities which share geographic, socio-demographic and historical characteristics.[14]

The geographic data (rainfall and temperatures, 1929-2010) are from the Meteorology Spanish National Agency; from the National Center of Geographic Information (elevation data); and daily river flow data from 1929 to 2010 from the Centre of Studies and Experimentation of Public Infrastructures (CEDEX from its acronyms in Spanish).

Population data comes from the Spanish Population Census (1920-2010) from the Spanish National Statistics Institute (INE). Demographic characteristics (gender, education, and civil status) are from the self-digitised 1940's Spanish Population Census; and the publicly available 1991's and 2011's Population Census (gender, education, civil status, age, and nationality).

[14] Spain is divided into different regions (NUTS 2). A region is split into provinces (NUTS 3), a province into districts or *comarcas* and a district into municipalities (LAU 2). The traditional territorial division of *comarca* is legally defined in Aragón, Cataluña, País Vasco (75% of the Ebro) and historically defined in other regions.

Based on Tur-Prats and Valencia-Caicedo 2020, I use geo-referenced information on the mass graves related to the Civil War and Franco's dictatorship (1936-1975) by the Spanish Ministry of Justice. This dataset includes data on mass graves' location, number of exhumed bodies, and political affiliation.

Table A.2.3 in the Appendix displays the descriptive statistics for the main variables.

2.4 Empirical Framework

This paper aims to estimate the long-term effects of hosting internally displaced population on social participation outcomes over the 50 years after the arrival of the displaced population. In this section, I describe the treated and control municipalities, the baseline specification and how I identify the causal effects.

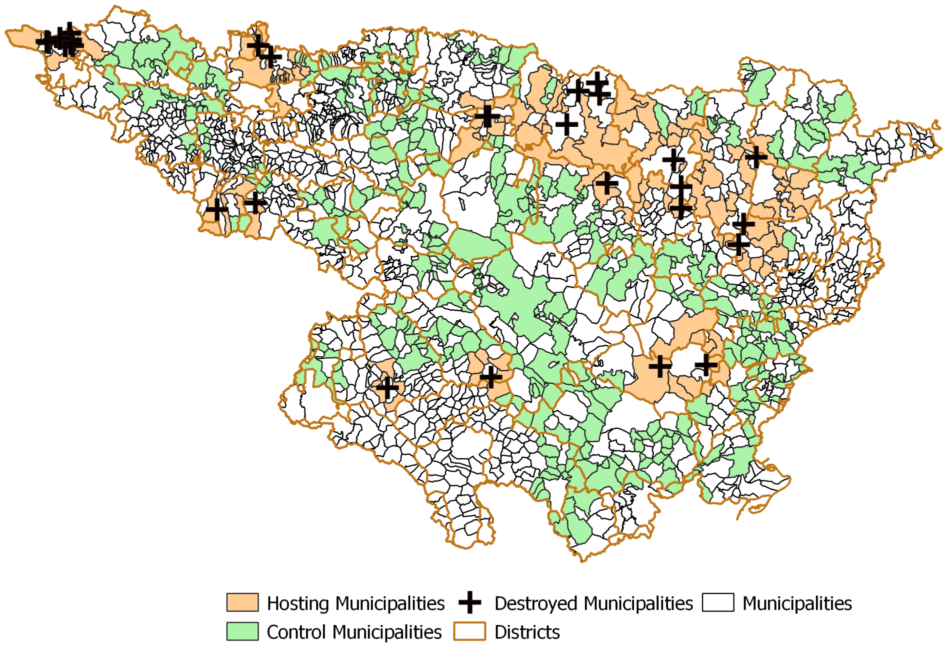
2.4.1 Treatment and Counterfactual

Treatment: *Host* municipalities. I am interested in capturing the effect of being a municipality that *host* the population who arrives induced by the destruction of villages due to reservoirs built during the dictatorship. In line with historical evidence, *host* municipalities are the bordering municipalities to the municipalities demolished by a reservoir. Hence, I define the treatment by the proximity to the demolished municipalities by these constructions.

First, I identify the municipalities destroyed by reservoirs and then pinpoint their neighbouring municipalities. To do so, I compute the exact location of destroyed municipalities by matching demolished municipalities with modern geo-referenced locations. See Figure A.2.4 for a treatment design and Figure 2.4 for its spatial distribution.

We could be concerned with the definition of my treatment. For instance, the affected population could have relocated within the same municipality of the reservoir, only to one or a subset of surrounding municipalities, or to other municipalities further away. I rely on descriptive evidence and the literature on migrants' destinations to validate my treatment definition. I show that while surrounded municipalities to destroyed municipalities experience an increase in their aggregate population after the year of displacement, destroyed municipalities have a decrease in their aggregate population. I use the changes in population between the population census before and after the year of displacement to calculate the changes in their aggregate population. Figure A.2.5 plots the population changes for the example of the Santa Ana reservoir (in Huesca and Lérida provinces). The descriptive evidence responds to previous work by Saldarriaga and Hua 2019 and Calderon-Mejia and Ibañez 2016

Figure 2.4: Treated and control municipalities



Note: This map illustrates the treated and control municipalities. Municipalities in orange correspond to the treated municipalities. Treated municipalities are the bordering municipalities to destroyed municipalities (black crosses). Lighter green municipalities correspond to the counterfactual: bordering municipalities to reservoir municipalities with no municipality destroyed. The orange lines represent the districts, and the black polygons the municipalities.

.[15]. Some of the changes in the population may be related to natural movements. I adjust the population changes by using the number of births and deaths per capita in a given province. The results are unchanged.[16] Additionally, I repeat the exercise of Figure A.2.5 by including the aggregate population changes in the surrounding municipalities to the bordering municipalities to a destroyed municipality in Figure A.2.6. Although a minority of the population could have relocated to other municipalities. Figure A.2.6 rules out the hypothesis that most of the population moves to municipalities further away.

Counterfactual induced by reservoir location. I design a counterfactual using the location of reservoirs. In particular, I exploit the presence of reservoirs which did not destroy any villages. This approach allows me to isolate the effects of internally displaced population inflows from the pure effect of the exposure to a reservoir nearby (Duflo and Pande 2007).

Municipalities bordering municipalities with one reservoir, built before the outcome year and did not destroy a village, function as my control group for *host* municipalities. It means that the counterfactual includes the bordering municipali-

[15] Saldarriaga and Hua 2019 and Calderon-Mejia and Ibañez 2016 find that destination municipalities closer to places of origin attract more displaced people.
[16] There are no historical data on births and deaths at the municipal level. The Figures can be provided upon request.

ties to any reservoir, unconditional if built before, during or after the dictatorship. Additionally, I further remove from my counterfactual the municipalities adjacent to villages which were also destroyed by reservoirs built after 1975, the end of the dictatorship. This strategy enables me to rule out the possibility that reservoir-induced displaced population inflows after 1975 could affect my results. Figure 2.4 illustrates the counterfactual.

Figure A.2.7 rules out the hypothesis that the control municipalities' population also increases. I do so by conducting the same exercise of Figure A.2.5 in the control municipalities. This result is particularly important because many of the reservoirs meant the arrival of new workers from other regions. Most of them were prisoners from concentration camps. The definition of the treatment and counterfactual overcomes potential co-founder effects. This phenomenon affects the construction of most reservoirs, without distinguishing whether a village was or was not destroyed.[17]

Sample. My sample is restricted to the treated and control municipalities. Treated municipalities are bordering on reservoirs built during the dictatorship, which destroyed one or more municipalities. Control municipalities are the bordering municipalities to a reservoir which did not destroy a municipality. I include in the counterfactual all the reservoirs built before the outcome year.

2.4.2 Baseline specification

Comparing bordering municipalities to reservoirs with and without population inflows, I estimate the long-term effects of hosting internally displaced population on social participation outcomes of municipality m in time t as follows:

$$Y_{mdpt} = \beta_0 + \beta_1 Host_{mdp(t-n)} + \beta_2 X_m + \alpha_d + \theta_{pt} + \epsilon_d \tag{2.1}$$

where Y_{mcpt} denotes social participation at time t (t= 1976-2015) for municipality m, district d, and province p. I measure social participation as voter turnout (in general and municipal elections), the number of new cooperatives created, new agricultural cooperatives established and new non-profit associations created in a given year. $Host_{mdp(t-n)}$ is equal to one if the municipality m hosted population at time *(t-n)*, zero otherwise. n is the number of years from the arrival of displaced population to time t. [18]

[17] The data on the location of camps come from *Los colonos de la España Verde de Franco's project.* Table A.2.4 shows that the results hold when controlling for the presence of concentration camps.

[18] Regions (NUTS 2 of the European standardize nomenclature) is the first administrative level in Spain, provinces (NUTS 3) the second, districts (or *comarca*) the third, and municipalities the fourth (LAU 2).

X_m is a vector for pre-treatment characteristics of municipality m. It includes demographic characteristics on gender, marriage, and education from the 1940 Population Census (the earliest census available before the massive construction of reservoirs during the dictatorship).

I also control for district fixed effects, α_d, and province-year fixed effects, $\theta_p t$. A district or *comarca* is a group of municipalities which share geographic, sociodemographic and historical characteristics. Therefore, district-fixed effects control for time-invariant characteristics that could affect the social participation of every municipality within a district. Province-year interactions account for annual shocks common across districts in a province. The regions have fiscal and political autonomy, but each province is differently affected by an economic or social shock due to their sociodemographic composition and geographical features.[19] To leverage the effects, I only exploit cross-municipality treatment variation in a district for identification. Finally, the social participation residual is likely to be correlated for municipalities in the same district. So, I cluster the standard errors by district, ϵ_d.

2.4.3 Identification: Instrumental variable strategy

Although the designed counterfactual enables me to isolate potential effects that derive from the pure impact of a reservoir (Duflo and Pande 2007), endogeneity problems occur when the explanatory variable may correlate with potential omitted variables related to the location of reservoirs, destruction of a village, and selection into a destination. To address these challenges, I implement an instrumental variable strategy by combining three sources of variation.

First, the location of reservoirs may not be random. Demographic or political factors could have affected the reservoirs' location, encapsulating potential endogenous problems. Hence, OLS estimates can be biased. I exploit the margin of whether a reservoir was planned before the dictatorship to incorporate an exogenous force in the location of a reservoir.

Second, I am not interested in evaluating the impact of reservoirs on social participation but the effect of migration driven by the destruction of municipalities by these constructions. Therefore, I need a plausibly exogenous variation in the probability of being a destroyed municipality. I show evidence that the likelihood of a municipality's destruction directly depends on the size of the reservoir. Then, I use the size of a reservoir to estimate the probability of being destroyed.

Third, although the proximity of a municipality to a reservoir defines the treatment and control, a certain level of self-selection into the destination still exists.

[19] The results hold when I control for region-year fixed effect instead. See Table A.2.20 .

To upfront this limitation, I use the marginal variation in the distance between a municipality and the closest reservoir.

Exogeneity in reservoirs' location: A pre-dictatorship plan. In 1933, the Second Spanish Republic designed the Water Infrastructures National Plan, which projected the reservoirs built in Spain during the following decades. Namely, the 1933 Plan proposed 32% of the reservoirs built during the dictatorship.

To overcome potential endogeneity problems in the location of reservoirs, I use the 1933's Plan, a pre-dictatorship water infrastructure plan. In particular, I exploit the margin of whether the closest reservoir was planned or not in 1933's Plan, in other words, before the dictatorship. This approach is very in line with the current literature, which has adopted historical infrastructure plans for the design of identification strategies (Duranton and Turner 2011).

Still, we could be a bit sceptical about how the 1933's Plan pinpointed the reservoir location. The Plan explicitly describes the set of geographical-climatic factors that determine the suitability of a reservoir: (i) Soil, (ii) Rainfall, (iii) Temperature, (iv) Altitude, and (v) River flow. I use data on average annual rainfall, average annual temperature, and altitude to empirically validate that these factors matter. Column (1) of Table A.2.5 shows that rainfall and altitude increase and decrease the probability of having a reservoir. On Table A.2.5 I generate further evidence suggesting that politically motivated decisions did not affect the decision to build a reservoir.

Instrumenting Village Destruction: Reservoir's size. As a second step in the identification strategy, I rely on an additional source of variation to measure the probability of a village being destroyed. I do so by using the size of the closest reservoir (reservoir's area in squared meters). As in Sarsons 2015 and Duflo and Pande 2007, I construct a measure of the normalized deviation of reservoir-area from its average.[20]

Intuitively, larger reservoirs (that expropriate higher land extensions) are more likely to destroy a village and, as a result, force its population to be relocated to a new place. Hence, bordering municipalities to a larger reservoir are more likely to receive population relocated by a reservoir. Figure A.2.8 supports this intuition by illustrating the positive correlation between the size of a reservoir and the probability of destroying a village.

The extension of each reservoir was also considered in the 1933 Plan. Depending on the characteristics of the valley, if it is wide and open, the floodplains can occupy densely populated areas. In other cases, especially in high and rugged areas, the reservoir occupies little populated or uninhabited land. Using climatic data on

[20] Normalized deviations of reservoir-area from its average level are calculated as $(Variable_{mcp}$-$AveVariable_{ebro})/SdVariable_{ebro}$

altitude, average annual rainfall, and average annual temperature, Table A.2.5 shows that larger reservoirs are more likely to be built in locations with lower altitudes and, consequently, with lower average annual rainfall and temperature. Ideally, data on the extension of a valley would be a perfect measure. Nonetheless, this data are not available.

Selection into Destination: Distance to Reservoirs. Ideally, we could randomize the location where the population affected ends up. However, this is far from reality and a common threat in the migration literature (S. Becker and Ferrara 2019). Although the counterfactual partially addresses this threat, marginal self-selection into the destination still exists. Therefore, as a last step in the identification strategy's design, I use the distance from municipality m to the closest reservoir.

One of the limitations of this paper is that I can not pin down the specific inflow in each destination. The best I can do is to identify the population living in a destroyed municipality before the destruction and assign these numbers to their bordering municipalities as the estimated inflow. This implies the important assumption that all surrounding municipalities are affected equally. However, extensive literature shows how the distance between the origin and destination locations strongly affects the amount of forcibly displaced populations, my treatment intensity. I rely on this literature (Rozo and Vargas 2021; Depetris-Chauvin and Santos 2018; Calderon-Mejia and Ibañez 2016; Alix-Garcia and Saah 2010) to proxy the amount of forcibly displaced population by using the distance from the centroid of each bordering municipality to the closest reservoir.[21] I measure the normalized distance deviation from its average in line with the reservoir's size variable. For my analysis, I restrict the sample to the bordering municipalities to the destroyed municipalities. However, to generate some evidence on the relationship between the intensity of forcibly displaced population and the distance to the closest reservoir, I extend the sample to the surrounding municipalities to the *host* municipalities. Figure A.2.9 shows that the closer municipalities are, the higher the likelihood of hosting more population.

As a last step in building my instrument, I combined the three sources of variation from the closest reservoir: a pre-dictatorship plan, reservoir size and distance. In particular, I interact the margin of whether the closest reservoir to municipality m was planned or not in the 1933 Plan to the area of the closest reservoir weighted by the inverse of the distance from the municipality m to the closest reservoir (Calderon-Mejia and Ibañez 2016).[22]

[21] The results hold when I use distance to the border of each municipality.

[22] Calderon-Mejia and Ibañez 2016instrument the inflow of internally displaced population with the cumulative number of massacres in the city c at time t, weighted by the inverse of the distance from the site of the massacre and the c^{th} city.

$$\text{IV} = Plan1933_{mdp} * \frac{Area_{mdp}}{Distance_{mdp}}$$

where $Plan1933_{mdp}$, is a dummy variable equal to one if the reservoir was projected in the 1933 Plan, zero otherwise. $Area_{mcp}$ is the extension in meters squared of the closest reservoir, and $Distance_{mdp}$ is the distance in meters to the closest reservoir.

The functional form of the instrument suggests that the probability of municipality m of being a *Host* municipality will increase with the size of the closest reservoir but decrease with the distance from the closest reservoir. Thus, the municipalities most affected by the arrival of forcibly displaced population would be those most geographically proximate to larger reservoirs.

The "first-stage" regression in the IV strategy is then given by

$$Host_{mdp(t-n)} = \delta_d + \phi_1[Plan1933_{mdp} * \frac{Area_{mdp}}{Distance_{mdp}}] + \beta_2 X_m + \alpha_c + \theta_{pt} + \eta_d \quad (2.2)$$

$Host_{mdp(t-n)}$ is equal to one if the municipality m hosted population at time *(t-n)*, zero otherwise. The regression includes district-fixed effects, year-province fixed effects, and controls for pre-treatment sociodemographic characteristics. My identification strategy, therefore, relies on within-district differences in the likelihood of being a host municipality.

I report the first stage at the bottom of Table 2.1. When a reservoir was projected before the dictatorship for each one-unit deviation in the ratio $\frac{Area_{mdp}}{Distance_{mdp}}$ from its mean, the likelihood of being a host municipality fall by 0.39 percentage points. The effect is significant at the 1% level, but I check for the instrument's strength using the F-statistic. I have an F-statistic of 50, which suggests that our instrument is strong enough for identification.

An alternative approach would be restricting my sample to reservoirs planned before the dictatorship and comparing treated and control municipalities. It would imply a comparison between municipalities surrounding a destroyed municipality during the dictatorship versus those bordering a municipality with a planned reservoir (built or not). There are three main empirical limitations to this approach. First, overcoming the endogenous problem related to the location of reservoirs is not enough. I need a plausibly exogenous variation in the probability of being destroyed. Second, a certain level of self-selection may still exist when deciding where to end up. Third, I would not be able to disentangle the main effect from the pure impact of a reservoir. I could remove from the sample the surrounding municipalities to a non-constructed reservoir to overcome the last threat. However, I would face a new limitation: substantial sample size and variation reduction. This paper relies on a within-district analysis. Only 8 out of the 107 districts in my sample have more than one reservoir planned in 1933 and constructed afterwards. The sample would decrease by 60%. Table A.2.6 produces the analysis using the alternative approach. Only the voter turnout in the municipal elections is statistically significant and becomes positive.

2.5 Results

In this section, I present evidence of the long-term effect of hosting forcibly displaced populations on decreasing social participation in hosting communities, as specified in equation (1). In the first half of Table 2.1, I show the OLS and IV estimates in the second half of Table 2.1. The rest of this section discusses the results by the outcome of interest: voter turnout in general and municipal elections, the total number of cooperatives per capita, the total number of agricultural cooperatives per capita, and the number of non-profit associations established yearly per capita.

Table 2.1: Effect of being a *host* municipality on social participation

VARIABLES	(1) Gen. Elec. Turnout	(2) Mun. Elec. Turnout	(3) Cooperatives	(4) Agrarian Coop.	(5) Associations
PANEL A: OLS estimates					
Host	-0.007122**	-0.002892	-0.000013	0.000015	-0.000369***
	(0.003205)	(0.006124)	(0.000031)	(0.000013)	(0.000122)
Observations	6,907	5,053	12,256	12,256	12,256
PANEL B: IV estimates					
Host	-0.146086***	-0.125955*	-0.000046	-0.000000	-0.006705***
	(0.053703)	(0.071344)	(0.000045)	(0.000012)	(0.001598)
Observations	6,907	5,053	12,256	12,256	12,256
PANEL C: First stage					
VARIABLES	Host	Host			
$Plan1933_{mdp} * \frac{Area_{mdp}}{Distance_{mdp}}$	-0.004606***	-0.004424***			
	(0.000640)	(0.000628)			
F-statistic	51.83	49.60			
Observations	6,907	6,907			
Outcome mean in 1977-2019	0.75	0.75	0.000062	0.000025	0.000552
Number of districts	85	85	85	85	85

Note: This table shows the effect of being a *host* municipality (between 1939 and 1975) on social participation from 1977 to 2019. I measure social participation as "voter turnout in general elections" (column (1)), "voter turnout in municipal elections" (column (2)), "number of cooperatives created yearly" (column (3)), "number of cooperatives in the agricultural created yearly" (column (4)), and "number of associations created yearly" (column (5)). I regress social participation on a dummy variable on the treatment: 1 if municipality m was a *Host* municipality during the Spanish dictatorship (1939-1975), 0 otherwise. I control for pre-treatment municipality characteristics (share female, share illiterate, and share single in 1940), district fixed effects and province-year interactions fixed effects. Standard errors are clustered at the district level. In panel A, I produce the OLS estimates. Panel B documents the IV estimates with the following instrument: interaction between the margin of whether the closest reservoir to municipality m was planned or not in the 1933 Plan to the area of the closest reservoir weighted by the inverse of the distance to the closest reservoir. Finally, panel C shows the first stage. The dataset is at the municipality and year level (for the years of elections for the turnout outcomes). Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Turnout in General Elections. Column (1) of Table 2.1 shows that being a *host* municipality is associated with a decrease in voter turnout in the general elections from 1977 to 2019.

Panel A presents the OLS estimates. Being a *host* municipality reduces general elections' voter turnout during the democracy (1977-2019) by roughly 0.7 percentage points with respect to the municipalities adjacent to municipalities that have a reservoir, but they were never destroyed. The effect is statistically significant at the 1% level.

While the IV estimates of Panel B are also negative, the point estimates are substantially larger than the OLS estimates. *Hosting* forcibly displaced population decreases voter turnout in general elections by 15 percentage points with respect to the municipalities adjacent to municipalities with a reservoir that did not destroy a municipality (the control group). The effect is significant at the 1% level and sizable in economic magnitude, corresponding to a 20% decrease in the baseline mean, indicating that *host* municipality considerably reduces their political participation in the general elections.

Turnout in Municipal Elections. Column (2) of Table 2.1 presents the estimates on municipal elections voter turnout. Being a *host* municipality also decreased the voter turnout in local elections from 1977 to 2019.

Although the OLS estimates of Panel A are not statistically significant, we can observe that the direction of the estimates is negative. Estimates that align with the IV estimates of Panel B. Being a *host* municipality reduces voter turnout in municipal elections by 13 percentage points with respect to the counterfactual. The effect is significant at the 10% level and sizable in economic magnitude, corresponding to a 17% decrease in the baseline mean. The results suggest that *host* municipalities, although also get less politically engaged in the municipal election, their reaction is slightly smaller.

Total Cooperatives. I look at the number of cooperatives per capita created from 1977 to 2018 at the year level. Column (3) shows that the OLS and the IV estimates are negative but statistically non-significant. There are no effects when I restrict the outcome to cooperatives in the agriculture sector. Column (4) shows the statistically non-significant estimates. The magnitude of the coefficient for agricultural cooperatives is even much smaller.

We could expect that the number of cooperatives per capita created in a year would decrease, given a sudden increase in its population after the arrival of the forcibly displaced population. I repeat the analysis of the number of cooperatives with two alternative outcomes: the absolute number of cooperatives in a municipality relative to its population and the number of cooperatives created in a year relative to the native population at the moment of the treatment. The results of Columns (1), (2), (4), and (5) of Table A.2.7 indicate that indeed there were no changes in cooperatives.

Total Associations. Column (5) of Table 2.1 reports the long-term effect of hosting forcibly displaced population on the number of new non-profit associations created in a year relative to its population from 1977 to 2018.

In line with voter turnout results, the OLS estimates of Panel A show that *host* municipalities are 0.04 points less likely to create an additional association than those adjacent to municipalities with a reservoir that did not destroy a village (significant at the 1% level).

The IV estimates are also statistically significant at the 1% level, but the magnitude of the effect is larger. *Host* municipalities are 0.7 less likely to have one more association per capita from 1977 to 2018 with respect their control group. The average number of associations in the municipalities in my sample is three, and less than one is the average number of new associations in a year. Therefore, the magnitude of the effect is considerable.

I conduct the same exercise as with the number of cooperatives. I also observe a decrease in the number of associations with an outcome on the absolute number of associations in a municipality relative to its population and an outcome with the number of associations created in a year relative to the native population at the moment of the treatment. Columns (3) and (6) of Table A.2.7 show the results which help to rule out the hypothesis that the decrease in the association's outcome is driven by the increase in the population instead by a net decrease in the associations in the *host* communities.

2.5.1 Interpretation of results

The findings indicate an overall decrease in social participation. I evaluate the effects of five different measures of social participation, which look at social participation from different angles.

Participation in local elections clearly indicates local social ties and involvement in the community, whereas other motives may drive participation in national elections. Therefore, the decrease in the political engagement in the municipal elections is a sign of a reduction in the community engagement of the citizens in the *host* communities.

Participation in the general elections accounts for the political engagement of citizens. A decline in the voter turnout in the general elections in the *host* municipalities could be interpreted as a decrease in institutional trust, lack of interest, or a form of protesting against a particular institution or regime.

An association implies the social gathering of individuals with a common interest or purpose. Hence, changes in the number of non-profit associations mirror a change in the social cohesion within communities. In contrast, a change in cooperatives may capture the volatility of the local economy. Reducing the number of new associations in the *host* municipalities could mean the fragmentation of the community and a decrease in inter-group social cohesion.

The no effect on cooperatives may echo the no differences in the economic impacts between treated and control municipalities, where their local economy may be equally affected by the construction of a reservoir.

Does the intensity of the treatment affect the results?. To answer this question, I add an interaction to the total forcibly displaced population relative to the native population already living in the *host* municipality at the moment of the arrival. Table 2.2 produces the results. The interaction is positive and statistically significant. One additional forcibly displaced individual increases general and municipal voter turnout by 3 and 2 percentage points, respectively (significant at one and 10% in Columns (1) and (2)). I observe similar results in the number of associations in Column (5). One additional forcibly displaced individual means an increment of 0.11% in the likelihood of having an additional association.

The effects correspond to a 20, 15 and 15% increase in general elections participation, municipal elections participation and number of associations, respectively. Although the net effects remain negative, these results suggest that the treatment intensity mitigates the impacts on social participation. The higher the forcibly displaced population who arrive in a municipality, the smaller the reduction in social participation.

Table 2.2: Effect of treatment intensity on social participation

VARIABLES	(1) Gen. Elec. Turnout	(2) Mun. Elec. Turnout	(3) Cooperatives	(4) Agrarian Coop.	(5) Associations
Host	-0.134955***	-0.117440**	-0.000048	0.000001	-0.006276***
	(0.039798)	(0.059022)	(0.000044)	(0.000012)	(0.001092)
Host * $\frac{IDP}{native}$	0.028176***	0.024014*	-0.000005	0.000002	0.001076***
	(0.008812)	(0.013410)	(0.000009)	(0.000004)	(0.000316)
Observations	6,907	5,053	12,256	12,256	12,256
Outcome mean in 1977-2019	0.75	0.75	0.000062	0.000025	0.000552
Number of districts	85	85	85	85	85

Note: This table shows the IV estimates of the effect of the intensity of the treatment on being a *host* municipality (between 1939 and 1975) on social participation from 1977 to 2019. I measure social participation as "voter turnout in general elections" (column (1)), "voter turnout in municipal elections" (column (2)), "number of cooperatives created yearly" (column (3)), "number of cooperatives in the agricultural created yearly" (column (4)), and "number of associations created yearly" (column (5)). I use the number of forcibly displaced population affected by a reservoir relative to the native population to measure treatment intensity. I regress social participation on a dummy variable on the treatment (1 if municipality m was a *Host* municipality during the Spanish dictatorship (1939-1975), 0 otherwise) and an interaction to the treatment intensity ($\frac{IDP}{native}$). I control for pre-treatment municipality characteristics (share female, share illiterate, and share single in 1940), district fixed effects and province-year interactions fixed effects. I instrument the treatment variable, *Host*, with an interaction between the margin of whether the closest reservoir to municipality m was planned or not in the 1933 Plan to the area of the closest reservoir weighted by the inverse of the distance to the closest reservoir. Standard errors are clustered at the district level. The dataset is at the municipality and year level (for the years of elections for the turnout outcomes). Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

I measure the treatment intensity using the total population in the destroyed municipalities from the population census before the destruction or treatment year. I may capture population changes driven by other factors, such as new births and deaths in a year. I adjust my treatment intensity measure by calculating the total births and deaths per capita. This information is at the province and year level and comes from the Spanish Statistical Office. I explain the measurement adjustment in the data appendix. The results of Table 2.2 hold. Table A.2.8 presents the results.

When did the effects start?. To learn about the persistence of the effects, I use the data on social participation available before 1977: municipal-level data on the number of cooperatives and associations created every year from 1945 to 1975.[23] The treatment years go from 1936 to 1962. Therefore, I restrict the outcome of interest to the years after the treatment year to conduct the IV analysis.[24]

I show in Table 2.3, Column (3), that the negative impacts on the number of associations started since the treatment year and persisted during and after the dictatorship for more than 50 years. A short-term effect of *hosting* forcibly displaced population exists on the decrease in the number of non-profit associations per capita created from 1945 to 1975. *Host* municipalities are 0.4 percentage points less likely to have a new association with respect to the counterfactual (significant at the 1% level). Similarly, the treatment intensity mitigates the likelihood of an additional association at 0.06%. There are no effects on new cooperatives created (Columns (1) and (2)). I provide further evidence of the persistence of the results by evaluating the lag effects on the number of new associations from the treatment. The results hold during the 50 years after the treatment. See Figure A.2.10.[25] When estimating the lag effects on the number of additional cooperatives created yearly, there is no effect. See Figure A.2.11

The results on when and for how long the results persisted can only be drawn for the number of cooperatives and associations. And, I can not make conclusions about these exercises to voter turnout since no elections took place during the dictatorship.

2.5.2 Threats

Balance sample. An alternative explanation for my finding is that individuals in *Host* municipalities initially had lower levels of social participation than the counterfactual and that these lower levels of social participation continue to persist today. I use sociodemographic municipal-level data from the Population Census of 1940 and social participation data in 1945. Table A.2.9 shows that that is not the case. Additionally, the sample is balanced between *host* and *control* municipalities on pre-treatment sociodemographic characteristics that could affect social participation.

Political drivers behind a destruction. Even if many reservoirs were already projected during the Second Republic, we could be concerned that a certain political margin could exist in deciding whether to build a reservoir. I benefit from data on the

[23] Table A.2.2 shows that the sample is balanced between *host* and *non-host* municipalities in the number of associations and agricultural cooperatives in 1945. There is no data available before 1945.

[24] For instance, if the year when the displaced population arrived at the municipality m was 1952 (treatment year), I restrict the sample to the years between 1952 and 1975.

[25] This Figure plots the IV coefficients of different regressions of equation (1) restricting the sample to n years after the arrival of the displaced population. n goes from 5 to 50 years.

Table 2.3: Effect of being a *host* municipality on social participation (1945-1975)

VARIABLES	(1) Cooperatives	(2) Agrarian Coop.	(3) Associations
PANEL A: All municipalities - reservoir in dictatorship			
Host	-0.000023	-0.000003	-0.003791***
	(0.000020)	(0.000005)	(0.000933)
PANEL B: Municipalities with a planned reservoir in 1933			
Host	-0.000024	-0.000002	-0.003540***
	(0.000020)	(0.000005)	(0.000634)
Host x $\frac{IDP}{native}$	-0.000002	0.000001	0.000633***
	(0.000004)	(0.000002)	(0.000191)
Observations	28,064	28,064	28,064
Outcome mean in 1977-2019	0.000062	0.000025	0.000552
Number of districts	85	85	85

Note: This table shows the IV estimates of *Host* municipalities (between 1939 and 1975) on social participation from 1945 to 1975. I measure social participation as "number of cooperatives created yearly" (column (1)), "number of cooperatives in the agricultural created yearly" (column (2)), and "number of associations created yearly" (column (3)). I regress social participation on a dummy variable on the treatment: 1 if municipality m was a *Host* municipality during the Spanish dictatorship (1939-1975), 0 otherwise. I control for pre-treatment municipality characteristics (share female, share illiterate, and share single in 1940), district fixed effects and province-year interactions fixed effects. Standard errors are clustered at the district level. In panel A, I produce the baseline estimates. Panel B shows the estimates with the intensity of the treatment (total forcibly displaced population affected by a reservoir relative to the native population). I can not repeat this exercise for turnout outcomes. No elections took place during the dictatorship The dataset is at the municipality and year level (for the years of elections for the turnout outcomes). Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

intensity of conflict during the dictatorship to proxy political-motivated decisions. [26] To measure the intensity of conflict, I use the geo-referenced information from the Spanish Ministry of Justice on the location of mass graves related to the Civil War and Franco's dictatorship.

I show, in Table A.2.5, a positive correlation between having a mass grave and a reservoir in a given municipality, as well as with the size of a reservoir. The intensity of the conflict is negatively correlated to the size of a reservoir. Most mass graves are located outside cemeteries (62% at the national level), typically alongside roads. There may be geographical-climatic factors in common between the location of reservoirs and common graves. The concern would be if Franco used the destruction of villages as a political weapon. To test this hypothesis, I repeat the same correlation exercise on a sample of reservoirs planned in 1933 (before the conflict started) and a sample built during the democracy. Panels B and C show the exact relationship between the location of reservoirs and conflict. These results make me think that we should not be worried about the possibility of political-motivated factors in deciding whether to build a reservoir in a particular municipality.

Exclusion restriction. The critical issue in an instrumental variable approach is whether the instrument satisfies the necessary exclusion restriction. That is,

[26] The political persecution used during the dictatorship was institutionalized. The largest share of the repression took place in small municipalities with less than 4,000 inhabitants, often mixing personal grievances with political causes (Arnabat Mata [2013]).

having close enough a reservoir planned in the 1933 Plan, its size and distance to the reservoir are correlated with factors other than receiving population from destroyed municipalities that may affect how voter turnout and the number of associations is today.

The most apparent reason why the exclusion restrictions may not be satisfied arises because the size of a reservoir is correlated with the historical rainfall in a municipality, and this, in turn, is positively correlated with current rainfall in nearby municipalities, which is negatively correlated with voter turnout (Gomez, Hansford, and Krause 2007). For this reason, in my IV estimates, where we use the size of a reservoir as an instrument, I test if the estimates hold when I also control for current rainfall. Table A.2.10 shows the very similar IV estimates.

As is generally the case with instruments, it is possible that despite my second-stage controls, my instrument still does not satisfy the necessary exclusion restriction. I undertake several falsification tests to provide some sense of my instrument's validity. My identification assumes that the arrival of the forcibly displaced population is the only channel through which the instrument affects social participation. Therefore, if my identification assumption is satisfied, I should not observe a similar negative relationship between the instrument (or any of their variables) and social participation in the parts of the Ebro region where the destruction of villages did not occur.

These are precisely the falsification exercises that I undertake. Specifically, I regress my social participation outcomes on my instrument for a sample of municipalities non-bordering a municipality destroyed by a reservoir. The OLS estimates in Table A.2.10 are statistically non-significant.

OLS vs 2SLS estimates. The OLS and the 2SLS are not very different from each other. However, they do not estimate the same thing. The 2SLS estimates a local average treatment effect (LATE), whereas OLS estimates an average treatment effect (ATE) if the error term is uncorrelated to the outcome. The LATE is the effect of being a host municipality (treatment) on social participation measures (outcome) for those observations that were induced to take up treatment by the instrument, $Plan1933_{mcp} * \frac{Area_{mcp}}{Distance_{mcp}}$.

Any comparison between OLS and 2SLS has to mention the caveat that the two estimates apply to different populations. The result could be very different if there is endogeneity in the model, being the OLS estimates biased and/or underestimated.

Intention to treatment. Which sub-population-group is driving these results? Are the natives or the newcomers who drive the results? Ideally, information at the individual level would allow me to shed some light. However, I cannot identify the sub-population group that is driving these results with the data I have, being the main caveat of this paper. My data only allows me to obtain an intention to treatment (ITT) effect. But, I can use data on the estimated inflow of population that arrived into the *host* municipality to generate suggestive evidence on the compositional effect. I have followed this approach when evaluating the intensity of the treatment. In the next section, I profit from contemporary data on the year of birth to look at the heterogeneity of the results when the age composition of the population changes in *host* municipalities.

2.5.3 Heterogeneity

Table A.2.11 explores the heterogeneity of the main results by adding interaction to three variables:

In Panel A, I start by looking at the total population share who experience the traumatic arrival of the population still living at the *host* municipalities. I call them *survivors* because they were the newcomers or the native. To conduct this analysis, I profit from the population information by age from the Population Census from 1991 to 2011. The primary assumption to calculate the total number of *survivors* is that the individuals have always lived in a given municipality. Table A.2.11, in Panel A, shows similar results to the estimates when I include a variable on the treatment intensity (Table 2.2). A higher share of *survivors* in the *host* municipalities mitigates the effect on contemporary social participation. The heterogeneous results suggest that over time native and forcibly displaced populations converge into a common understanding and a sense of belonging, which could translate the hosting effect into an increase in political and community engagement.

I can also look at whether there is heterogeneity in the effect by the decade of the treatment. The timing of the treatment goes from 1936 to 1962. Panel B shows the results. I can observe in Column (1) that the net effect of an arrival of a forcibly displaced population on voter turnout in the general elections during the 1930s and 1960s is smaller compared to the treatment in the 1950s. These findings could be extended to the effects on the number of associations, including the decade of the 40s. In the 1950s, the number of constructed reservoirs peaked, which revolved in many municipalities being destroyed by these constructions (See Figure A.2.12 for a descriptive presentation). This feature could explain the heterogeneous results when the treatment took place. Additionally, Table A.2.12 produces the results by successively dropping the treated municipalities in the 1930s and 1940s. The results become statistically non-significant when dropping the treated municipalities in the 1950s.

Panel C of Table A.2.11 also shows the differential effect by the decade in the outcome (1977-2019). The results suggest that the decrease in social participation was softer in the 1980s, 1990s, 2000s and 2010s compared to the 1970S. Notably, a sequential decrease in the voter turnout in national elections over time with respect to the 70s exists. The decrease in participation in the municipal elections and the number of new associations is relatively stable over time.

2.6 Robustness Checks

In this section, I present evidence of the validity of my results in four ways. First, I test the validity of my results by estimating some falsification tests. Second, I examine the existence of possible confounding factors. Third, I approach my question using an alternative sample and treatment definition. Finally, I include robustness checks with different fixed effects, clusters, and covariates.

Falsification Tests. The identification of this paper relies on the plausibly exogenous variation across municipalities in the probability of receiving a large inflow of forcibly displaced population. I conduct two sets of falsification tests to rule out the possibility of hidden bias.

First, suppose the arrival of forcibly displaced population (treatment) cause the decrease in turnout and number of associations. In that case, we should expect no causal relationship on my outcomes before the treatment. In Table A.2.13, I run two different falsification tests in this direction using data before the treatment. In Panel A, I look at the impact of hosting forcibly displaced population during the 1950s-1970s on cooperatives and associations in the year strictly preceding the treatment. There are no differences between treated and control municipalities in the number of new cooperatives and associations. I find the same results (In Panel B of Table A.2.13). Reservoirs constructed during the democracy (1976-2018) also destroyed some municipalities. I repeat the analysis with reservoirs constructed after 1977 and look at the potential effects of outcomes in 1977. The results are statistically non-significant. The economic development reservoirs generate could have mitigated the effects.

Second, a share of the planned reservoirs in 1933 was never built. I profit from this information to generate a placebo counterfactual. I compare municipalities bordering a reservoir which received population forcibly displaced by this construction against bordering municipalities to a municipality where a reservoir was projected but never built. We should expect to observe an effect, but the estimates would also capture the effect of a reservoir. Panel C of Table A.2.13 produce the results, which are also negative and statistically significant, but more minor than the baseline estimates.

These exercises eliminate the hidden effect hypothesis and bring confidence in interpreting the IV estimates of Table 2.1 as causal.

Confounding Factors. The main concern with my setting is the potential confounding effect of the violence perpetrated during the Spanish Civil War and the dictatorship (1936-1975) on modern-day social participation. During the dictatorship, those who opposed Franco were persecuted, tortured, and many of them murdered. Previous literature has documented that the Spanish Civil War had a long-lasting effect on the decrease in social capital and voting behaviour (Tur-Prats and Valencia-Caicedo 2020). In line with the existing literature, Panel A of Table A.2.14 shows how the intensity of violence during the Civil War and the dictatorship decrease social participation. Given the potential endogeneity of conflict, we should interpret the OLS estimates carefully.

Nevertheless, do the main effects change conditional on the same level of conflict? Table A.2.14 shows that the main results hold when I include the level of violence fixed effects (Panel B) and drop the municipalities with a mass grave (Panel C). The drop in the same size could explain the soft decrease in the magnitude of the effects. Only 16% of the Ebro catchment area municipalities have a mass grave. Moreover, 13 out of the 18 provinces in the Ebro catchment area have mass graves.

Ebro's catchment area overlaps with eight regions, remarkably diverse in economic development and population density. We could be worried about potential cofounders related to the local labour market. To assess potential regional differences, I remove region by region in order of lowest development in the main specification.[27] I document in Table A.2.15 that the effects do not change when I drop the four poorest regions in my sample (Castilla-La Mancha, Comunidad Valenciana, Cantabria, and Castilla y Leon, which correspond to a 16% of the sample). However, when I continue removing sequentially more regions, voter turnout becomes insignificant (when dropping 25%) and associations (when dropping 64%). While province-year interaction accounts for annual shocks standard across counties in a province, it does not allow me to control for local-specific labour market characteristics. To overcome the potential confounding effect, I control for modern-day labour market data at the municipal level. I measure the local labour market with data on the share of the unemployed population, the share of the employed population in the agricultural sector, and rural depopulation in the 50's.[28] The effects on turnout and the number of associations established remain unchanged. The estimates on municipal participation in elections are negative but statistically non-significant. I present the result in Table A.2.16.

Alternative Treatment Definition. In this paper, I focus on the effects of being a *host* municipality of the population displaced by reservoirs constructed dur-

[27] I classify each region using GDP per capita at the regional level for 2018.

[28] There was a significant out-migration flow during the 50s in Spain from backward regions to the leading areas, with rural families heading for urban areas (Pinilla and Sáez 2016). I use population changes between the Population Census in 1950 and 1960 to capture this variation.

ing the dictatorship. Still, four cases of displacement happened in the Ebro region during the democracy. I use these events to evaluate if the effects hold when the treatment occurs after the dictatorship. Table A.2.17 documents the OLS and IV estimates. It is unsurprising to see statistically non-significant results given the small variation generated by only four reservoirs.[29] The voter turnout in the general and municipal elections remains negative, but the outcome on associations is now positive. The destruction of a municipality by this infrastructures produced a political disappointment. Citizens hoped that the end of the dictatorship would come with the end of these events. A larger decrease in institutional trust could explain the increase in the magnitude of the coefficients. Additionally, it was common to observe a higher social cohesion between affected and non-affected populations, positively impacting general trust.

The data for the cooperative and association outcomes are for every year from 1977 to 2018, whereas turnout data are only available for the years of elections. Table A.2.18 shows that the results of Table 2.1 are unchanged when keeping the years of my sample constant across outcomes.

I define the treatment, *host* municipalities, as those strictly bordering the contemporary municipality destroyed by a reservoir built by Franco. In Table A.2.19, I relax this definition. In Panel A, I extend the treatment to the surrounding municipalities to the treated or *host* municipalities. I call them *extended host* municipalities. The estimates are consistent with the main results, with a little increase in the magnitude of the effects. However, the results do not hold when I restrict the treatment definition to the *extended host* municipalities (in Panel C). The central assumption of this paper is that most of the affected population settle down in the bordering municipalities. The no effect in a sample of *extended host* municipalities and the control municipalities validate this assumption. Additionally, Panel C documents the no effects when including the municipalities destroyed in the sample, providing additional evidence of the central assumption.

Alternative Specifications. A primary concern with the validity of my estimates is the potential correlation in error terms across space between different counties in Spain. I consider clustering my standard errors over a larger geographical level, i.e., at the province and region levels, respectively. In Table A.2.20, I replicate the IV estimates from Table 2.1 and report the corresponding standard errors and p-values under the different clustering assumptions. The cooperatives' outcomes are also statistically significant when I cluster the error terms at the province or re-

[29] The sample decreases compared to Table 2.1. Although the counterfactual is the same, the treatment now only includes the bordering municipalities to the four reservoirs which destroyed a municipality during the democracy. Only 61 municipalities are defined as *Host municipalities*. Control municipalities are the bordering municipalities to a reservoir which did not destroy a municipality. I include all the reservoirs built before the outcome year.

gional level. Moreover, the number of associations created per capita is statistically non-significant with a regional cluster. There are essential differences between the municipalities within a region, which can affect these results. However, the spatial differences between the unit of analysis can be mitigated by clustering the error terms at the district level (a group of municipalities with common geographical and historical characteristics).

When the impact of hosting forcibly displaced population is identified from between regions and between-years variation in the treatment, all the estimates are robust. To this end, I replicate the IV estimates from Table 2.1 by including province and year-region fixed effects. I present the results in Panel C of Table A.2.20. Turnout results do not change either when I control only for the region and time-fixed effects (in Panel D). The number of new cooperatives is now significant, and associations non-significant. Some of the regions in Ebro's catchment area are pretty vast. Many demographic and social unobservable characteristics may exist between municipalities within the same region, which could explain the differences in the results. Panel E shows that the main results hold when controlling for post-treatment socio-demographic characteristics (older than 64, women, educated, foreign population shares) in equation (1). I also show in Figure A.2.13 that the point estimates are very similar with and without controls and clusters.

2.7 Mechanisms

The main results show significant effects of hosting internally displaced population on the decrease in voter turnout in general and municipal elections and the number of new associations in the *Host* communities. What mechanisms are primarily responsible for driving these effects?

Reservoirs impacted natives and forcibly displaced populations differently. In this vein, I propose two main mechanisms that could explain the results. First, the perceived effectiveness of the construction of reservoirs within local communities led to changes in institutional trust and, consequently, affected citizens' political participation. Second, the conflict between the citizens positively and negatively affected by the same reservoir fostered distrust between the two groups within the *Host* communities, impacting their social cohesion.

A group of studies have shown that effective policymaking can lead to an increase in trust in the government (Acemoglu, Cheema, et al. 2020; Fair et al. 2017; León-Ciliotta, Zejcirovic, and Fernandez 2023), and as a consequence, affect citizen's political participation (Putnam 1995; Hetherington 1998). In previous research, Carlin.R, Love.G, and Zechmeister.E 2014 show that there is a decrease in trust in places where governments respond poorly to earthquakes in El Salvador, Haiti and

Chile. Nevertheless, the correlation reverses signs among those who feel the government response was adequate. Flückiger, Ludwig, and Önder 2019 also support these findings, where they find that the trust in central government increased disproportionately in regions that experienced a relatively large influx of relief-effort-related resources against the Ebola epidemic in West Africa. León-Ciliotta, Zejcirovic, and Fernandez 2023 find that the poor execution of public policies or malpractices in the processes also undermines citizen trust in their institutions in Peru. Therefore, the effect on institutional trust depends on the efficacy of government response in a particular location.

In recent years, some studies have found that experience of wars or other conflicts in society generates distrust (Kijewski and Freitag 2018; Rohner, Thoenig, and Zilibotti 2013; Fehr 2009; Nunn and Wantchekon 2011), while others have emphasised how war or violent experience can foster pro-social behaviour (for a detailed review, see Bauer et al. 2016). The contradictory results may, to a large extent, be associated with the nature of the violence experienced. In general, if the trauma occurred due to intra-group conflict, trust in others is reduced as' betrayal aversion' comes into effect (Booth et al. 2022; Fehr 2009). On the other hand, if the trauma was brought about by inter-group conflict, the within-group bond strengthens, and individuals become more cooperative within that group; however, the inter-group trust decreases. Bauer et al. 2016 finds that people more exposed to the conflict tend to increase their social participation by joining more local social and civic groups in their communities. Notably, violence affects in-group pro-social behaviour: participation with one's own village or identity group members.

Both of these mechanisms could be at play in my setting. Reservoirs are public goods investments and a considerable population benefits from these reservoirs' water and energy services. Nonetheless, their development comes at the cost of relocating thousands of families into new communities. The perception of the efficacy of government response may depend on the group citizens belongs: *winners* or *losers*. The mechanism of perception of efficiency in the government response would imply that I should observe an increase in the trust in the government itself in the *Host* communities with a higher share of the native population relative to the affected population by a destroyed municipality. I may expect the opposite effect in those communities with a higher share of the affected population by a destroyed municipality versus the native population at the moment of the treatment. Second, water constructions have detonated considerable tensions at the local level, dividing the population between those in favour (*winners*) and against a construction (*losers*). We could consider the native population among the *winners* and the citizens directly affected by a destroyed village among the *losers*. The arrival of the population from neighbouring destroyed villages by these infrastructures accentuated the inter-group tensions. The hypothesis of inter-group conflicts would thus imply a decrease in trust in other people.

In the following subsections, I empirically test the relevance of these potential mechanisms.

2.7.1 Efficiency in the construction of reservoirs

How a reservoir impacts each citizen may affect their satisfaction with constructing a reservoir and, consequently, their perception of government intervention. The majority of the population in the Ebro region was involved in agriculture, and the quality of irrigation systems was essential to household welfare. Whereas the population who gained the benefits may feel that the government response was adequate, the citizens who lost their properties and communities may consider the policy poorly implemented. Hence, effective policymaking may lead to an increase in institutional trust.

The native population have got along with the internally displaced population in the *Host* municipalities. Therefore, the direction of the effects would depend on the relative share of the forcibly displaced population. In most cases, the number of the forcibly displaced population who arrived at *Host* municipalities did not exceed the native population. So, I should observe an increase in institutional trust in the government responsible for the reservoir in the *Host* municipalities. Nonetheless, the sign of the effects would decrease when including the intensity of the treatment. By including interaction to the relative weight of the affected population by a destroyed municipality relative to the native population in the year of arrival, I can proxy the composition of the *Host* municipality. Ideally, I would like to identify the displaced and native populations explicitly. But the data does not allow this identification.

I obtain data on institutional mistrust from the Spanish Sociological Research Center (CIS, Centro de Investigaciones Sociológicas) between 1989 and 2015. In Table 2.4, I test the above-presented hypothesis by replicating equation 1 with an outcome on institutional mistrust. The institutional mistrust outcome is equal to the share of the respondent which did not vote in the general election for institutional mistrust reasons. Panel A of Table 2.4 show the OLS and IV results. Column (1) shows that the OLS estimate is statistically non-significant. Column (3) produces the IV estimate, which is positive and statistically significant at the 5% level. Being a *Host* municipality increases the mistrust of the democratic institutions by 42% compared to the counterfactual. Furthermore, the composition of the *Host* municipality matters. The treatment intensity, measured as the relative weight of the internally displaced population, decreased institutional mistrust by 13%. Panel B of Table 2.4 presents the results.

Table 2.4: Effect of being a *host* municipality on trust

VARIABLES	(1) mistrust institution	(2) general trust	(3) mistrust institution	(4) general trust
PANEL A: OLS and IV estimates				
Host	-0.021128	3.154025***	0.416575**	32.123407***
	(0.026713)	(0.791905)	(0.198143)	(8.773012)
PANEL B: With treatment intensity				
Host	0.029027	-3.105458***	0.473722**	44.875916***
	(0.031396)	(1.056824)	(0.193811)	(12.246582)
Host x $\frac{IDP}{native}$	-0.038112***	7.940735***	-0.133221***	-23.914819***
	(0.012657)	(0.969884)	(0.042808)	(8.252557)
PANEL C: With *survivors* share				
Host	-0.005706	3.246849***	0.374143**	33.562357***
	(0.027502)	(0.808511)	(0.164589)	(7.710376)
Host * *survivors* share	-0.100313**	-0.754467	-0.249174***	-10.350062***
	(0.043911)	(1.296813)	(0.077116)	(3.348593)
Observations	1,288	546	1,288	546
Number of districts	47	27	47	27

Note: This table produces the effect of being a *host* municipality (between 1939 and 1975) on trust from 1989 to 2015. I measure trust as "institutional mistrust" (columns (1) and (3)), and "general trust" (column (2) and (4)). I proxy institutional trust with the average share of individuals who reported mistrust in the national government for each municipality, and general trust with the average general trust. Trust data come from the Spanish Sociological Research Center (CIS, Centro de Investigaciones Sociológicas). I regress trust on a dummy variable on the treatment: 1 if municipality m was a *host* municipality during the Spanish dictatorship (1939-1975), 0 otherwise. I control for pre-treatment municipality characteristics (share female, share illiterate, and share single in 1940), district fixed effects and province-year interactions fixed effects. Standard errors are clustered at the district level. In panel A, I produce the OLS (columns (1) and (2)) and IV (columns (3) and (4)) estimates. In Panel B, I interact the treatment to the treatment intensity (number of forcibly displaced population affected by a reservoir relative to the native population to measure treatment intensity). Finally, Panel C shows the results with an interaction to the total population share born before the arrival of population that still live at the *host* municipalities (*survivors*),. The dataset is at the municipality and year level (for the years of elections for the turnout outcomes). Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

The interpretation of the results of Table 2.4 should be taken with a grain of salt. The data on mistrust I use may capture the mistrust in the democratic institutions, which could differ from the trust in Franco's government. I do not have data on trust in the institution in charge of the reservoirs (Franco's government). Nonetheless, recent evidence suggests that voters may punish and reward politicians in the next elections based on demonstrated performance in managing certain events. For example, Fowler and Hall 2018 find that leaders are punished for droughts, floods, and even shark attacks that occur under their watch. Maffioli 2021; Cole, Healy, and Werker 2012 confirm that voters reward governments that increase disaster spending in response to the Ebola epidemic in Liberia and extreme rainfall in India. So, incumbents fare better when they respond to a crisis with emergency relief. To validate this hypothesis, I then use data on voting share for Falangist parties from 1977 to 2019 as a proxy of support or satisfaction with Franco's government.[30] On one hand, I show that being a *host* municipality increases the vote share for Falangist parties in the general elections by 0.1%. This result is statistically significant at

[30] Falangist parties are far-right parties aligned with Franco's ideology. Falangism (Spanish: falangismo) was the political ideology of two political parties in Spain that were known as the *Falange*, namely first the *Falange Española de las Juntas de Ofensiva Nacional Sindicalista* (FE de las JONS) and afterwards the *Falange Española Tradicionalista y de las Juntas de Ofensiva Nacional Sindicalista* (FET y de las JONS).

10% and sizable, corresponding to a three times higher baseline mean from 1977 to 2019. See the results on Panel A of Table A.2.21. On the other hand, a higher share of the internally displaced population compared to the native population in the treatment year decreased the vote share for Falangist parties by 0.01% (Panel B). The estimates on the voter turnout in municipal elections are statistically non-significant, suggesting that political engagement at the municipal level is a sign of community engagement rather than institutional trust.

Which is the interpretation of these results? The arrival of the forcibly displaced population seems to accentuate the spectrum between the benefits the *winners* gained and the cost the *losers* paid. While the native population seems to have been satisfied with the performance of Franco's government, the internally displaced population seems upset with the ineffective policy implementation. The results suggest that the native population rewarded Franco's government in three ways: increasing their support for Falangist parties at the polls, which translates into a permanent mistrust in the democratically elected government and decreased political participation in the general elections during the democracy. On the opposite side of the spectrum, the internally displaced population has expressed its dissatisfaction and disappointment by taking action against the responsible government (Franco's government): decreasing their mistrust in the democratically elected government, which has led to a decrease in the support for Falangist parties at the polls, and an increase in their voter turnout in the general elections. All these effects mean that voters account for the effectiveness of government responses, and the voter gratitude and displeasure can persist decades later in their actions (Bechtel and Hainmueller 2011).

Even if the evidence presented is quite informative. Some things could be improved in the analysis above. First, I only look at the demographic composition in the treatment year. The composition between native and internally displaced populations could change over time. I do not have data to control for these changes. To generate some supportive evidence, I conduct the same analysis of Panel B of Table A.2.20 by adding an interaction to the contemporary share of the population born before the treatment year. I call them *survivors*. The intuition is that places where more internally displaced population arrived are more likely to have a higher share of survivors. Following this assumption, I show that the results hold. A higher share of survivors decreases mistrust in the democratic government (Panel C of Table A.2.20), and decreases Falangist's vote share (Panel C of Table A.2.21). Second, to validate the efficiency perception mechanisms, I should expect not to find the same results for the *Host* municipalities of the affected population during the democracy. I show in Table A.2.22 that it is the case. Finally, I generate empirical evidence on the correlation between institutional mistrust and voter turnout in how institutional mistrust affects voter turnout in general elections. I produce the results on Column (1) of Table A.2.23.

2.7.2 Inter-group clashes in *Host* municipalities

Historically, water construction has been a controversial project. Reservoirs are public goods that are constructed at the cost of many villages in rural areas. It is common to see two opposing groups in reservoir construction: those in favour (farmers and energy companies) and against construction (population affected by construction and environmentalists). The case of Riaño (a village in the province of León), which attracted nationwide attention, provides a vivid illustration. Franco started Riaño's reservoir in 1965, a reservoir projected before the dictatorship. The project stopped with the beginning of democracy, but it restarted in 1982. In May 1986, Riaño villagers launched large-scale protests against the reservoir construction after they received notifications of their forced evacuations a few months earlier. In parallel, many other protests took place in favour of the construction. In 1987 the reservoir was filled, flooding nine villages. During the dictatorship, raising such resistance to construct a reservoir was unthinkable. However, protests pro and against reservoirs also existed in Ebro's region. Figure A.2.14 illustrate two protests in favour and against the construction of reservoirs, which help us imagine the tensions within a municipality among the population pro and against a construction. The picture on the top illustrates the opening ceremony of Oliana Reservoir in 1959 in Lérida province. The second picture shows a group of citizens demonstrating against the construction of the Santa Ana reservoir in the Huesca and Lerida provinces in the 60s.

On the one hand, the arrival of the population from neighbouring villages destroyed by these infrastructures could have accentuated the clashes between the pro and against groups, leading to a decrease in inter-group trust. On the other hand, the within-group bond strengthened among the most damaged by the destruction (internally displaced population), and internally displaced individuals became more cooperative within that group. Participation in local elections indicates local social ties, and the number of non-profit associations shows social cohesion within communities. Thus, the between-group conflict could explain the decline in non-profit associations, participation in the municipal elections, and its increase when the relative size of the internally displaced population rises.

I use data on general trust to test this hypothesis. The information on general trust comes from the Spanish Sociological Research Center (CIS, Centro de Investigaciones Sociológicas) from 1989 to 2015. The question related to trust in other persons asked in the surveys is, "Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?". Answers range between 10 (most people can be trusted) and 0 (must be very careful). I calculate the municipal average level of trust in others to repeat the analysis of equation 1 with the new outcome. The trust in others question is explicitly related to *most people can be trusted*. Therefore, the data value would depend on the demographic composition of native and internally displaced populations in *Host* communities in the survey year.

I show in Column (4) of Table 2.4 (Panel A) that the trust in others increases in *Host* municipalities by 32% with respect to the control municipalities. The IV estimate is significant at the 10% level, but the OLS estimate of Column (2) is statistically non-significant. As mentioned in the above paragraph, it is critical to interpreting the results of which group is predominant in a given *Host* municipality. In the majority of cases, most of the population is native. As a result, the effects capture an increase in within-group or intra-group trust among the native population. In Panel B, I show that when the number of forcibly displaced population increases, there is a decrease in trust in others of 23%. The latest estimate measures the decrease in inter-group trust. Intuitively, when more population is affected by a destroyed municipality, the inter-group conflict increases, leading to a decrease in inter-group trust. So, the within-group trust among citizens whose communities were destroyed by the reservoir became more cohesive within that group.

Why would we expect that the results persist in the long-term? The traumatic experience changed the values that parents in the *Host* communities passed on to their children (G. Tabellini 2008; Nunn and Wantchekon 2011). Arroyo and Eth 1995 emphasise that the traumatic events witnessed directly (through personal traumatisation) or indirectly (through parents' reactions to traumatic events) during the initial years of life might produce enduring effects on general trust.

As I stress above, I only look at the demographic composition in the treatment year. I repeat the same validation exercise as in the effectiveness mechanism. First, a higher share of *survivors* declines the general trust. Second, I find that *host* municipalities of individuals from municipalities destroyed during the democracy have a decrease in general trust compared to their counterfactual. The destruction of municipalities in the democracy and the protest in favour and against the construction was massively covered in the media, which could negatively affect the intra-group trust. At the same time, the relative size of the internally displaced population increases the inter-group trust. Table A.2.22 produces the results. Finally, I generate empirical evidence on the correlation between general trust and social participation outcomes. I show that general trust increases voter turnout in general elections (Column (1)), agricultural cooperatives (Column (4)) and non-profit associations (Column (5)). Surprisingly, when I interact with the dummy treatment variable, the sign of the estimates flips, and there is no effect on agricultural cooperatives. This result indicates that the increase in within-group among the native population decreases voter turnout in both elections, suggesting that large doses of social engagement can decrease turnout by consuming time, exposing citizens to conflicting views, and providing an alternative route to fulfilling civic obligations (Atkinson and Fowler 2014). However, we should be careful with the interpretation of the results. Measures of social capital, such as general trust, are positively correlated with the turnout, and reverse causation and omitted variables may bias the results. The intra-group trust among natives declines the likelihood of a new association by 0.02%.

2.7.3 Other potential mechanisms

Two additional mechanisms could explain the results. However, there are no data available to test them empirically.

First, the trauma of being expelled from a village to be destroyed could be a potential mechanism. Exposure to a traumatic experience might generate psychological distress Kijewski and Freitag 2018 and lead to the formation of pessimistic beliefs about the trustworthiness of others. The trauma could affect as well the participation in elections. Marsh 2022 concludes that traumatic events decrease turnout in the next election through the mechanism of post-traumatic stress-demobilisation responses.

Second, their municipalities' destruction and subsequent movement into bordering places significantly impacted social network cohesion. Some households may have been able to coordinate with pre-displacement group members during the transition to living nearby. However, coordinating with friends or relatives was only sometimes feasible, impacting the network size. Thus the degree to which households lived with pre-displacement network members could affect social cohesion. The dispersion of their pre-displacement networks made the creation of new non-profit associations much harder.

2.8 Conclusion

The number of forcibly displaced persons keeps increasing worldwide. How forcibly displaced populations are integrated influences social participation in the long-term, with vast implications on the economic, education, and health of the populations living inside hosting communities.

I study how changes in exposure to internally displaced population inflows that happened in the past affected social participation in *host* municipalities during the next 50 years. I measure forced displacement as internal displacement associated with reservoirs during the Spanish dictatorship (1936-1975) at the Ebro's river catchment area. I profit from three sources of variation to implement an instrumental variable strategy. First, to overcome the non-randomness of the reservoir's location, I exploit the margin of whether a reservoir was planned or not before the dictatorship. Second, I use the reservoir size that strongly affects the probability of a village being destroyed. Third, I profit from marginal variation in the distance to the closest reservoir to upfront the potential self-selection into a destination. To this end, I rely on a newly-collected historical panel dataset on forced displacement and social participation.

I find that *host* municipalities of internally displaced persons have a long-term decrease in social participation from 1977 to 2019. *Hosting* forcibly displaced population decreases the voter turnout in the general and municipal elections compared to the municipalities adjacent to municipalities with a reservoir that did not destroy a village (the control group). *Host* municipalities are also less likely to have an additional association per capita than their control group. Significantly, the number of forcibly displaced population relative to the natives mitigates the impacts. A reservoir impacted natives and forcibly displaced populations differently, which explains the two mechanisms underlying my results. First, voters accounted for the effectiveness of the government responsible for the construction with two opposing behaviours: whereas the natives rewarded, the internally displaced population punished the government of Franco. Their gratitude and displeasure have persisted for decades in their mistrust of the democratically elected institutions and voting behaviour (Falangist's party support and turnout). Second, the conflict between the citizens positively and negatively affected by the same reservoir fostered distrust between the two groups within the *host* communities, impacting their social cohesion and persistence over time.

My results highlight how the arrival of a population from nearby locations whose communities have been destroyed shapes citizen trust in institutions and other people, with long-lasting effects on political participation and engagement in non-profit organizations. I observe that the negative impacts on the number of associations started from the arrival of the forcibly displaced population, and the effects survive after 50 years. My findings suggest that citizens who lose trust in other persons or institutions are unlikely to regain general confidence, translating into a permanent decrease in social participation.

Understanding the long-term consequences of forcibly displaced population in *host* communities is crucial to improving the design of interventions targeting the inter-group cohesion (Mousa 2020). Furthermore, my results have important implications for how forcibly displaced populations poorly integrate into new destinations. Neglecting to respond to forcibly displaced population's integration may end up hurting social participation, with long-lasting consequences over the following decades, and ultimately negatively impacting welfare.

Appendix Chapter 2

A Database construction

This section describes in further detail the database I construct for this project.

A.1 Displacement data

I gather information on municipalities destroyed by reservoir constructions in Ebro's region.

To identify the destroyed municipalities, I classify the reservoirs into two groups: those destroying and non-destroying municipalities. Then, I find out the affected municipalities. Among the later, I also include the municipalities expropriated due to reforestation targets (three in total).[31]

This information comes from qualitative information in text or excel format from regional and national organisations (The Commission of Population Affected by Large Reservoirs-COAGRET, Ecologist in Action and Desplazados.org), local and regional institutions (Ayuntamiento de Mansilla de la Sierra, Agencia Catalana del Agua, Ayuntamiento de Mequinenza, Pajares de Cameros website, Turismo Zaragoza), civil society groups (Fundación Cerezales, Calatayud.org, Janovas.org), community associations created after a village disruption by a reservoir (Asociación Río Aragón contra el recrecimiento del embalse de Yesa, Colectivo 7 Villas del Alto Najerilla, Despoblados Huesca website, Caminos de Barbastro website, Geografía Infinita, among others) and digitised old local newspapers (El Periódico de Aragón, El diario.es, El País, ABC Aragón, La Vanguardia Lleida, El Correo, Vive Campoo). Table A.2.1 describes the data sources for the 19 reservoirs which generated the destruction of one or more municipalities.

I include as follow a description of the main variables designed:

Displacement year. I benefit from information on the year of expropriation at the reservoir level to proxy the year of displacement. Pinning down to the specific year when the affected population was relocated to a new destination is unavailable. First, the population was displaced in a staggered way. Second, the year of displacement is not documented for every case. Therefore, I proxy the year of displacement to the expropriation year. From a legal standpoint, expropriation

[31] During the Spanish dictatorship, riverine plots, located on the reservoir bank, were expropriated to reduce erosion through a reforestation process (Daumas.M 1976).

(with the owner's name and expropriation year) should be published in the Gazette. Nonetheless, during the dictatorship, it was not always the case. Only 58% of the cases were published (11 over 19 reservoirs generating displacement). When the expropriation year is non-available, I assign them the average number of years between the expropriation and construction. Nine years are the average years lag from the expropriation announcement year to the last year of construction.

Estimated Population displaced. I estimate the population displaced by exploiting the displacement year and the population stock data from the population census (1900-2010). In particular, I use the population stock before and after the year of displacement to estimate the population forcibly displaced. This paper differentiates three excluding scenarios. First, the estimated displaced population is the population stock from the census before the displacement year, when a municipality is entirely destroyed by a reservoir (64% of the cases). Second, when a village is partially destroyed, I estimate the population displaced as the population change between the census before and after the displacement year. For instance, if the year of construction is 1971, the estimated year of displacement is 1962 (equal to 1971-9). So the estimated displaced population is the variation in the population between the 1960 census (existent census before the estimated displacement year, 1962) and the 1970 Census (the next census). Third, I repeat the same exercises described for the second case when a destroyed municipality is reconstructed outside the affected area (4 municipalities in total).

The main caveat of this approach is that I can not exclude that a share of the population moved before the population census before the displacement year, suggesting that my estimates on the forcibly displaced population may be underestimated. On top of that, the estimated numbers may capture population changes driven by other factors, such as new births and deaths in a given year. I adjust my treatment intensity measure by calculating the total births and deaths per capita from 1941 to 1975. This information is available at the year and province levels. The data comes from the Spanish Statistical Office. I calculate the total births and deaths per capita by dividing the total births and deaths by the population in a province. I then discount the population changes with the changes related to newborns and deaths, using the total births and deaths per capita.

Host municipalities. I know some of the municipalities of destinations where most of the population moves to. However, I do not have information on the total population that arrived at each municipality (called *host* municipalities, see section 4.1 for more details). Although this is a significant limitation. There is anecdotal and historical evidence that shows that most of the affected population by a destroyed municipality settle down in the surrounding municipalities. I list n Table A.2.1 the municipalities of the destination I know and their data sources.

Table A.2.1: Data sources: destroyed villages by reservoirs in Ebro region

(1) Reservoir	(2) Construction Year	(3) Destroyed Municipalities	(4) Displacement Year	(5) Some Destinations	(6) Data Sources
Búbal	1971	Búbal, Saqués, Polituara, El pueyo de Jaca	1966		COAGRET BOE
Canelles	1960	Caserras del Castillo, Fet, Finestras, Monfalcó, Soriana	1951		COAGRET
Ebro	1952	Mediano, La Magdalena, Quintanilla de Valdearroyo, Quintanilla de Bustamante, Arroyo, Llano, Villanueva, Renedo, Quintanamanil, Orzales, La Población, Corconte	1947		El.diario.es ViveCampo.es
Escales	1955	Aulet, Casterner de les Olles	1946		COAGRET despobladosenhuesca.com 1 2
González Lacasa	1962	Ortigosa (Los Molinos)	1953		larioja.com
El Grado	1969	Clamosa, Lapenilla Mipanas, Torreciudad Ligüerre de Cinca, Escanilla	1962		COAGRET, BOE despobladosenhuesca.com 1 2
Lanuza	1976	Lanuza, Sallent de Gállego	1965	Sallent de Gállego, Sabiñánigo Jaca, Huesca	COAGRET, BOE 1 2 3 ABC.es
		Tramacastilla de Tena			
Mansilla	1960	Mansilla de la Sierra	1951	Mansilla, Villavelayo	mansilla.org, BOE 1
Mediano	1959	Arasanz, Muro de Roda, Samitier, Gerbe y Griebal, Mediano, Ministirio, Morillo de Tou, Plampalacios	1950	La Fueva	COAGRET, BOE 1 2 lamaletavieja.com
Mequinenza	1964	Mequinenza	1951	New Mequinenza	yesano.com, ElMundo.es
Riba-Roja	1969	Fayón	1958	New Fayón	heraldo.es, BOE
Santa Ana	1961	Tragó, Boix	1962		lavanguardia.com
Las Torcas	1946				
La Tranquera	1959	Somed, Cocos Nuévalos, Carenas	1952	Cinco-Villas	BOE, elpais.com
Ullibarri-Gamboa	1956	Azúa, Garayo, Larrínzar Marieta, Mendíjur , Mendizábal Nanclares, Orenin, Zuazo Gamboa, Landa, Urizar Aroma, Esavarri	1947	Barrundia, Arrazua-Ubarrundia, Elburgo	elcorreo.com Ullibarri-Gamboa
Yesa	1959	Esco, Tiermas, Ruesta Bescós, Acín, Cenarbe Larrosa, Villanovilla Yosa, Bergosa	1948	Sigüés	BOE, Esco GeografíaInfinita
Jánovas	1960	Jánovas, Lacort, Javierre San Felices, Ligüerre Santa Olaria, Fiscal	1951		Elperiodico Elpais.com
N = 17					

Note: This table list the data sources to identify the municipalitites destroyed by reservoirs in Ebro region during the dictatorship (1936-1975)

Modern geo-referenced locations. I compute the exact location of municipalities by matching destroyed municipalities with modern geo-referenced locations.

Planned before the dictatorship. I extract from the 1933 Plan (or *Plan Nacional de Obras Hidraúlicas*) the reservoirs planned before the dictatorship. This information allows me to identify the reservoirs already planned before Franco among the reservoirs built in the dictatorship. 37% of reservoirs built during the dictatorship was already planned in the 1933 Plan. Table A.2.2 shows by period the total number of reservoirs generating municipalities destruction, planned reservoirs, and planned reservoirs which led to a municipalities destruction. I can also identify the location where the construction of a reservoir was rejected or understudy.

Additional information. I can identify which municipalities were entirely or partially destroyed by the construction of a reservoir. Very few of them were partially destroyed. Furthermore, my data allow me to identify which actions were implemented during municipality destruction. Some of the actions implemented were: the reconstruction of a municipality, partial reconstruction of a community, not entirely destroyed but neglected, the reservoir was never built, but the municipality was entirely expropriated.

Table A.2.2: Reservoir stock per period

	(1) Pre-dictatorship (before 1936)	(2) Dictatorship (1936-1975)	(3) Post-dictatorship (after 1975)	N
N	41	49	40	130
Destroyed	5	17	4	26
Planned	-	18	10	28
Destroyed and Planned	-	12	1	13

Note: This table shows the total number of reservoir by three main periods: pre-dictatorship (before 1936), dictatorship (before 1936), and post-dictatorship (after 1975). I also provide with the number of reservoirs destroying municipalities and the planned reservoirs. I define planned reservoirs as a reservoir that was planned in the 1933 Plan (or *Plan Nacional de Obras Hidraúlicas*). Source: Inventory of Dams and reservoir dataset. Spanish Ministry for Ecological Transition, data sources listed in Table A.2.1, and Plan Nacional de Obras Hidraúlicas.

A.2 Associations and Cooperative data

I also gather historical and modern-day data on the total number of associations and cooperatives from 1945 to 2018 at the municipal and yearly levels. This information comes from the Regional Registry of Associations (Registro Autonómico de Asociaciones) and Regional Registry of Cooperatives (Registro Autonómico de Cooperativas) of Aragón, Navarra, Cantabria, Cataluña, País Vasco, Castilla y León, La Rioja y Castilla la Mancha. It includes registration year, name, address, expiration year and sector.

Associations are defined as non-profit entities that are constituted by agreement of three or more legally constituted natural or legal persons, who undertake to share knowledge, means and activities to achieve lawful purposes of general or particular interest. There are specific associations that are registered on specific sectoral registries. For instance, political parties; unions; Business organisations; churches, denominations and religious communities; sports associations; consumer and user associations; and professional associations of Armed Forces members, the Civil Guard and magistrates, judges and prosecutors are among them. My data do not include the abovementioned associations.

Cooperatives are societies made up of people who join, under a free membership regime and voluntary withdrawal, to carry out business activities aimed at satisfying their economic and social needs and aspirations, with a democratic structure and functioning.

A.3 1940 Census data

I digitised the 1940 Spanish Population Census from the Spanish National Statistics Office to conduct this project. Figure A.1 shows a page image of the 1940 Spanish Population Census. Population data are disaggregated by gender, civil status and education (illiterate rate).

Figure A.2.1: 1940 Spanish Population Census printed

Censo de población de 1940 (Hecho)

Clasificación por municipios

Provincia de Huesca

MUNICIPIOS	VARONES				Alfabetos	Alfabetas	MUJERES			
	Total	S	C	V			V	C	S	Total
1 Abay	269	194	63	12	217	153	27	61	139	227
2 Abena	126	90	25	11	106	67	9	27	62	98
3 Abiego	276	150	106	20	193	212	46	122	134	302
4 Abizanda	193	108	72	13	141	118	15	72	91	178
5 Acín	113	76	28	9	105	70	17	33	53	103
6 Acumuer	169	111	48	10	141	100	24	48	86	158
7 Adahuesca	219	128	84	7	170	156	27	91	119	237
8 Aguas	121	71	43	7	95	81	24	45	55	124
9 Agüero	407	227	152	28	273	244	56	172	201	429
10 Aguinalíu	114	68	38	8	83	49	11	41	35	87
11 Aínsa	240	147	81	12	196	185	18	86	121	225
12 Aísa	134	77	51	6	111	65	19	51	55	125
13 Albalate de Cinca	614	346	227	41	426	440	71	237	301	609
14 Albalatillo	253	135	111	7	174	176	23	114	138	275
15 Albelda	621	300	288	33	498	517	81	320	279	680
16 Albella y Jánovas	439	259	151	29	358	267	45	151	196	392
17 Albero Alto	112	67	34	11	95	93	15	38	80	133
18 Albero Bajo	115	65	44	6	84	63	11	43	56	110
19 Alberuela de la Liena	120	72	39	9	89	71	18	43	59	120
20 Alberuela de Tubo	102	56	39	7	79	87	19	44	61	124
21 Alcalá de Gurrea	757	467	254	36	605	573	70	254	414	738
22 Alcalá del Obispo	162	94	62	6	119	106	33	69	79	181
23 Alcampel	990	466	472	52	737	678	123	487	411	1 021
24 Alcolea de Cinca	876	472	353	51	588	577	119	373	402	894
25 Alcubierre	549	296	228	25	444	431	67	240	283	590
26 Alerre	93	63	27	3	67	61	12	28	43	83
27 Alfántega	110	66	39	5	86	63	18	42	40	100
28 Alíns del Monte	40	21	16	3	25	27	1	19	26	46
29 Almudébar	1.558	938	563	57	1.055	970	176	568	738	1 482
30 Almunia de San Juan (La)	535	264	246	25	373	349	69	258	247	574
31 Almuniente	251	144	94	13	180	168	19	104	129	252
32 Alquézar	268	139	111	18	195	159	28	112	102	242
33 Altorricón	572	351	200	21	406	409	45	212	321	578
34 Angües	339	189	135	15	245	259	47	139	177	363
35 Aniés	238	130	89	19	165	123	20	92	124	236
36 Ansó	410	253	132	25	311	424	55	191	301	547
37 Antillón	168	89	68	11	128	141	25	76	97	198
38 Anzánigo	173	102	61	10	140	102	12	58	76	146
39 Apiés	205	131	64	10	150	141	38	60	98	196
40 Aquilué	133	86	44	3	112	68	11	39	55	105

— 3 —

Note: This image presents an example of one page of the Spanish 1940 Population Census. Population data are disaggregated by gender, civil status and education (illiterate rate). I digitised 18 provinces in total for this paper (Alava, Barcelona, Burgos, Cantabria, Castellón, Girona, Guadalajara, Guipuzcoa, Huesca, Lleida, Navarra, Palencia, La Rioja, Soria, Tarragona, Teruel, Vizcaya y Zaragoza. Source: Spanish National Institute of Statistics.

B Summary statistics

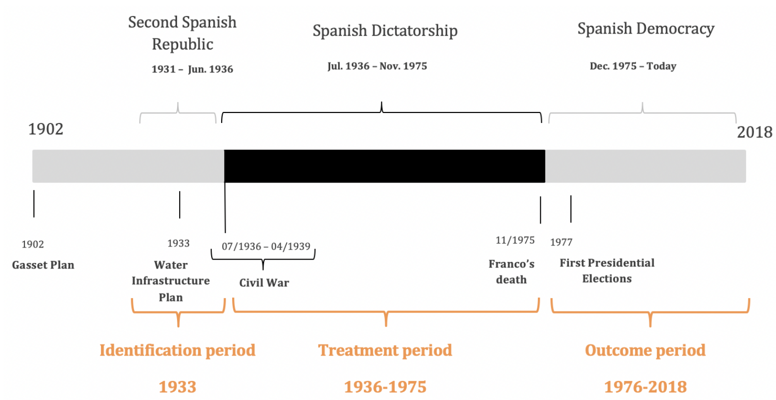
Table A.2.3: Descriptive Statistics

	mean	sd	min	max	count
Treatment Charac.					
Year Arrival Displaced Pop.	1949.74	6.74	1936	1962	6907.00
Displaced pop.	188.56	600.27	-773.00	5328.00	6907.00
Displaced vs native pop. (%)	0.26	0.88	-2.23	8.02	6907.00
Pre-treatment Charac. (in 1940)					
Population	2030.19	11139.69	66.00	238695.00	6907.00
Women Share (%)	0.49	0.08	0.00	0.72	6907.00
Literate Share (%)	0.70	0.14	0.00	1.00	6907.00
Married Share (%)	0.36	0.08	0.00	0.58	6907.00
Widowed (all) Share (%)	0.08	0.02	0.00	0.18	6907.00
Widowed (female) Share (%)	0.11	0.03	0.03	0.24	6907.00
Geographic Charac.					
Average annual rainfall, in $m^3, 1931-1932$	519.64	236.39	87.00	1346.46	6907.00
Average annual temperature, in Celsius, 1931-1932	101.03	27.85	7.50	167.63	6907.00
Altitude, in meter	578.60	280.84	26.00	1432.00	6907.00
Average annual river flow in $m^3/s1966-1975$	197.76	416.01	0.09	2240.90	1050.00
Reservoirs Number	1.27	1.22	0.00	5.00	6907.00
Social participation outcomes (1977-2019)					
General elections turnout	0.75	0.09	0.00	1.00	6907.00
Municipal elections turnout	0.75	0.13	0.00	1.00	5303.00
Cooperatives per capita	0.000062	0.000840	0.000000	0.076923	12256.00
Agricultural cooperatives per capita	0.000025	0.000360	0.000000	0.014011	12256.00
Associations per capita	0.000552	0.003541	0.000000	0.113514	12256.00
Demographic Charac., 1977-2019					
Women Share (%)	0.47	0.05	0.00	0.71	6907.00
Terciary Education Share (%)	0.07	0.06	0.00	0.35	6907.00
Above 64 Share (%)	0.50	0.25	0.00	0.93	6907.00
Foreigner Share (%)	0.06	0.08	0.00	0.63	6907.00
Other Charac.					
Number mass graves	0.61	1.09	0.00	9.00	6907.00
Total victims	15.21	182.65	0.00	4024.00	6907.00
Institutional mistrust	0.07	0.10	0.00	0.50	1415.00
Falangist parties vote share, 1977-2019	0.00	0.00	0.00	0.05	6907.00
General trust	2.33	2.31	0.00	7.50	588.00
N	6907				

Notes: This table reports descriptive statistics for the main variables and sample considered in the analysis. The analysis covers 1,695 municipalities (for the 14 general elections from 1977 to 2019). Pre-treatment characteristics are from the 1940 Population census. Municipality characteristics are from 1991 and 2011 Spanish Population Census.

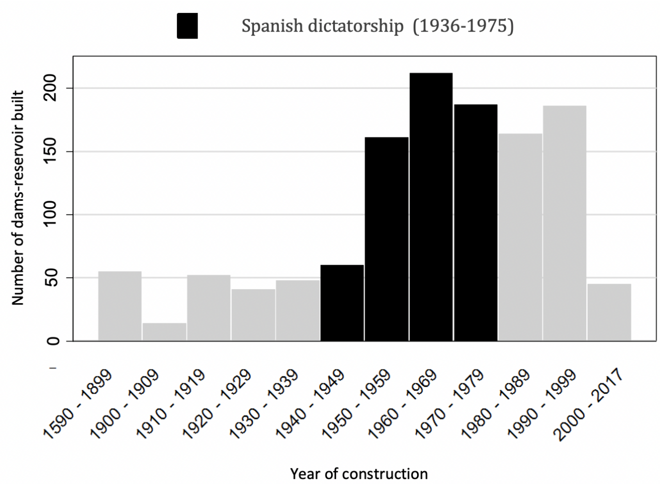
C Figures appendix

Figure A.2.2: Spanish Historical Timeline, 1900- 2018



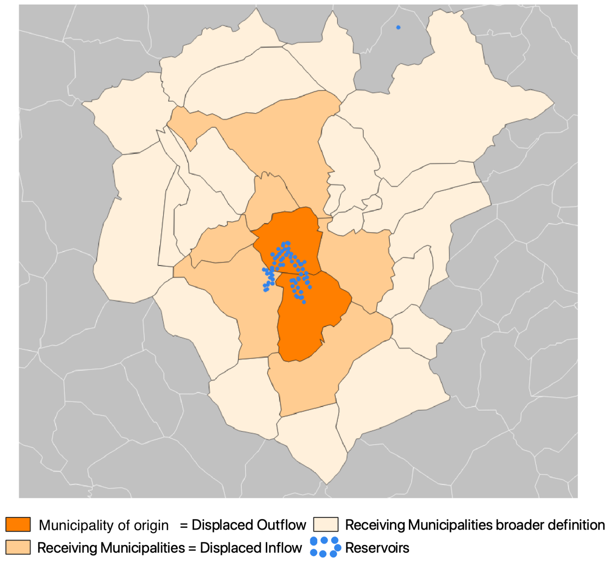
Note: This figure shows the historical timeline of Spain from 1900 to 2018. Darker colours correspond to the Spanish dictatorship (1936-1975). In orange, I illustrate the timing exploited by this paper.

Figure A.2.3: Construction of reservoirs in Spain



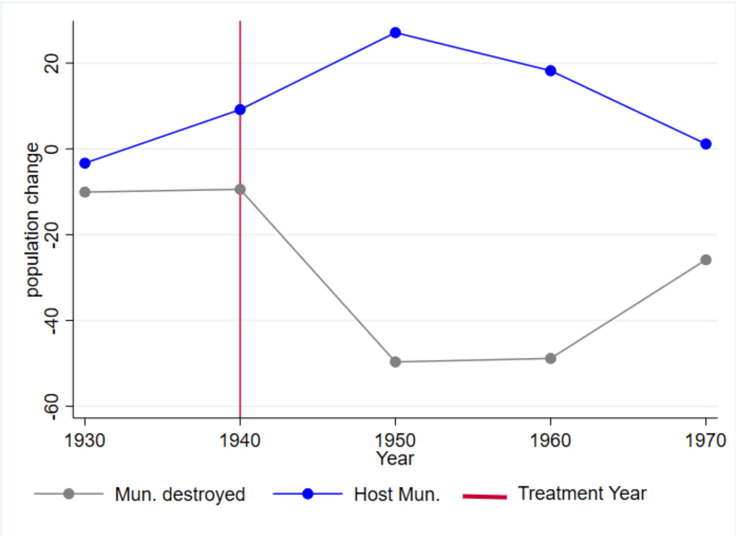
Note: This figure describes the number of reservoirs built from 1590 to 2017. Darker colours correspond to the Spanish dictatorship (after the Spanish Civil War (1936-1939). It shows a boost at the beginning of the Spanish dictatorship in 1940. Source: Inventory of Dams and reservoir dataset. Spanish Ministry for Ecological Transition.

Figure A.2.4: Spatial Treatments Construction



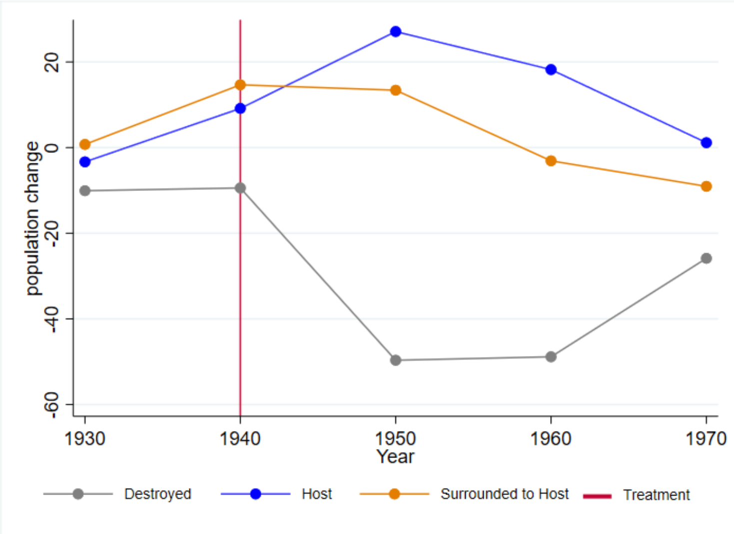
Note: This figure shows the treatment definition. Darker colours correspond to the destroyed municipalities or municipalities of origin. Medium intensity orange is the *host municipalities*, the bordering municipalities to destroyed municipalities. Lighter orange is the surrounding municipalities to the *host* municipalities. I call them *extended host* municipalities.

Figure A.2.5: Population changes in destroyed and host municipalities from the treatment year



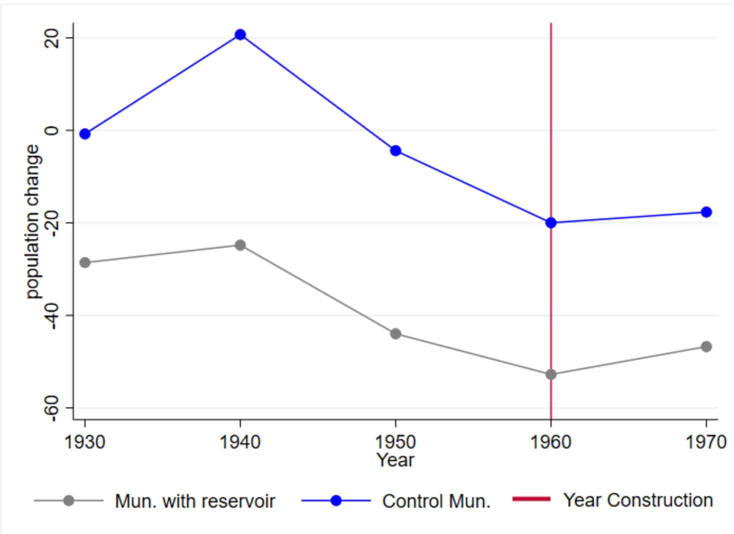
Note: This figure uses the example of Santa Ana Reservoir (in Huesca-Lleida Province) to show the aggregate population decrease in the origin municipalities and the population increase in the *host* municipalities after displacement. The grey colour line corresponds to the origin municipalities. And blue colours line to *host* municipalities. Source: Spanish Census 1920-1970

Figure A.2.6: Population changes in municipalities surrounding host municipalities



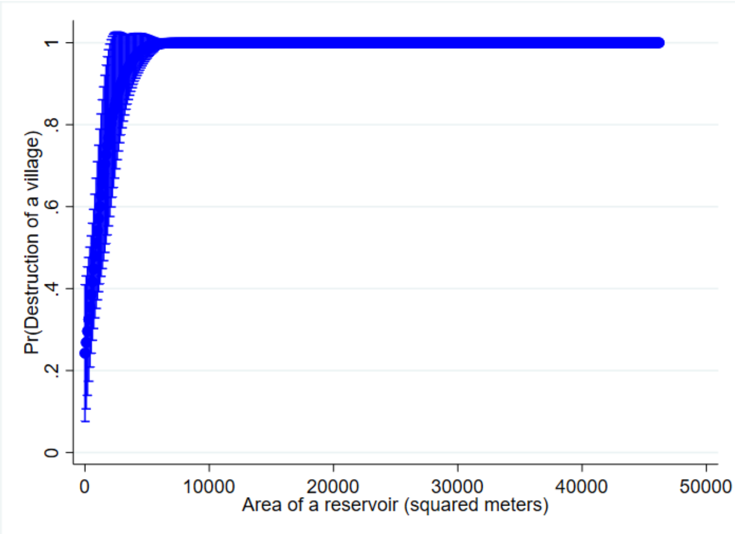
Note: This figure also uses the example of Santa Ana Reservoir (in Huesca-Lleida Province) to show that the aggregate population did not change in the municipalities surrounding the host municipalities (the orange line). Source: Spanish Census 1920-1970

Figure A.2.7: Population change in reservoir and control municipalities from the year of construction



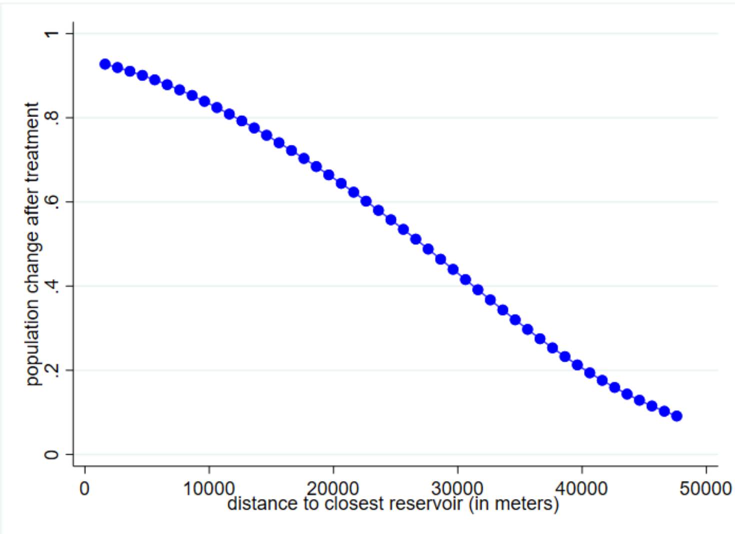
Note: This figure uses the example of Sotonera Reservoir (in HuescaProvince) to show that the reservoir and its surrounding municipalities follow the same patterns in the changes in their aggregate population. The grey colour line corresponds to the municipalities where the reservoir is built. And blue colours line with the surrounding municipalities. Source: Spanish Census 1920-1970

Figure A.2.8: Probability of a village being destroyed by area of a reservoir



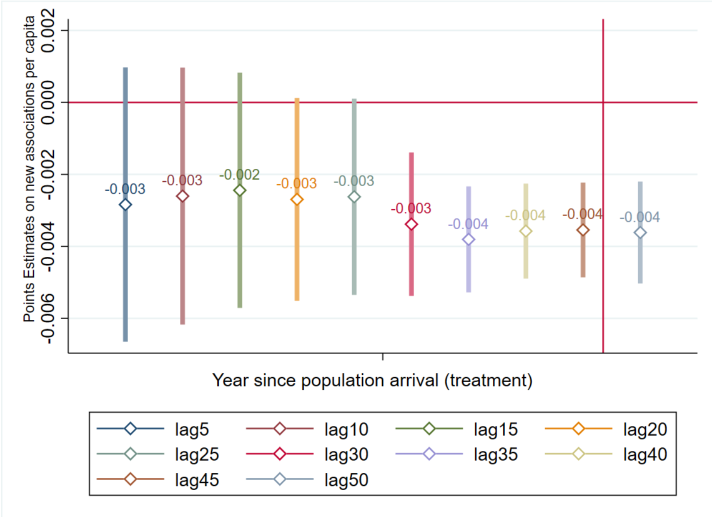
Note: This figure illustrates the relationship between the size of a reservoir and the probability of destroying a municipality. I measure the size of a reservoir with the area of a reservoir in squared meters from the Spanish Ministry for Ecological Transition. The larger the reservoir, the higher its probability of destroying a municipality.

Figure A.2.9: Forcibly displaced population and distance to a reservoir



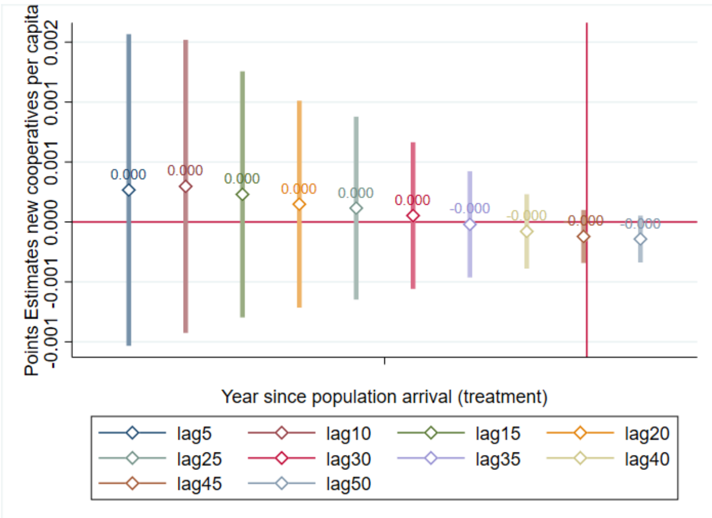
Note: This figure illustrates the relationship between the distance (in meters) from municipality m to the closest reservoir and the total population arriving at the municipality. To construct this figure, I measure the forcibly displaced population as the population change between the population census before and after the treatment year. I extend the sample to the surrounding municipalities to the *host* municipalities. Further, the municipality, the lower the population that arrives.

Figure A.2.10: Effects displaced population inflows on agricultural cooperatives (1945-2018)



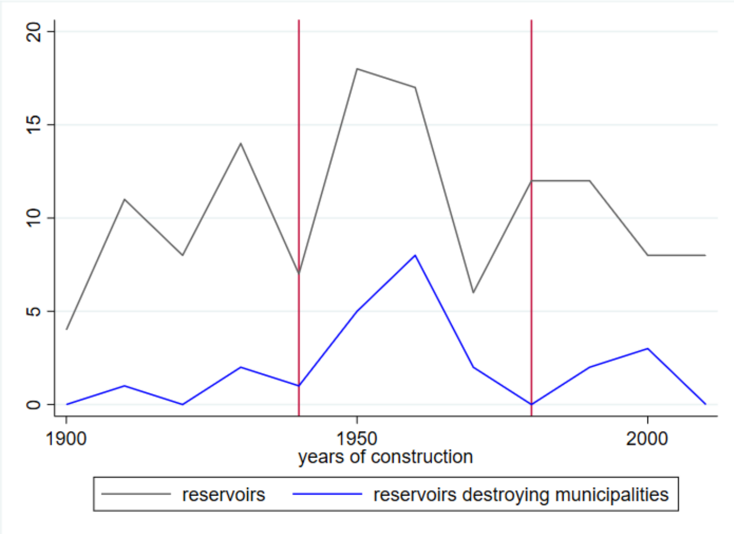
Note: This figure shows the effects of being a *host* municipality on the new associations by lag years. In the x-axis, I include the years from the treatment (the arrival of the forcibly displaced population). I include the point estimates in the y-axis. I repeat equation (1), extending the sample from 1945 to 2018. Treatment is defined at the yearly level.

Figure A.2.11: Effects displaced population inflows on agricultural cooperatives (1945-2018)



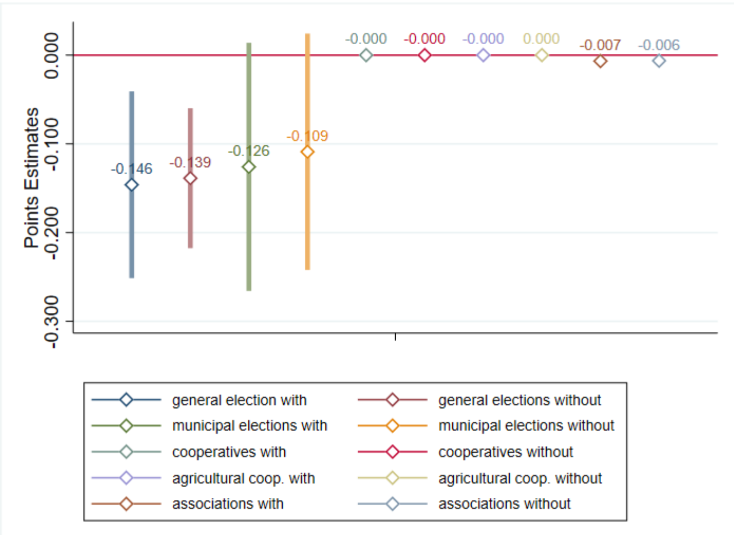
Note: This figure shows the effects of being a *host* municipality on the new cooperatives by lag years. In the x-axis, I include the years from the treatment (the arrival of the forcibly displaced population). I include the point estimates in the y-axis. I repeat equation (1), extending the sample from 1945 to 2018. Treatment is defined at the yearly level.

Figure A.2.12: Evolution number of reservoirs (1900-2017)



Note: This figure plots the number of reservoirs built in the Ebro region from 1900 to 2017. The grey line represents the total reservoirs constructed. This information comes from the Inventory of Dams and Reservoir dataset. Spanish Ministry for Ecological Transition. Blue lines are the number of reservoir-destroying municipalities. I describe the data sources in Table A.2.1.

Figure A.2.13: Point estimates with and without controls and clusters



Note: This figure shows the point estimates of equation (1) for the social participation outcomes with and without controls and clusters.

Figure A.2.14: Citizens pro and against reservoir constructions in Ebro region

Note: These images illustrate two protests in favour and against the construction of reservoirs, which help us imagine the tensions within a municipality among the population pro and against a construction. The first illustrates the opening ceremony of Oliana Reservoir in 1959 in Lérida province. The second picture shows a group of citizens demonstrating against the construction of the Santa Ana reservoir in the Huesca and Lerida provinces in the 60s. Source: José Demaría Vázquez and Juan Antonio Cemeli

D Complementary analysis appendix

Table A.2.4: Potential confounding effects of concentration camps

VARIABLES	(1) Gen. Elec. Turnout	(2) Mun. Elec. Turnout	(3) Cooperatives	(4) Agrarian Coop.	(5) Associations
PANEL A: With interaction to concentration camp distance					
Host	-0.143675***	-0.123624**	-0.000046	0.000000	-0.006594***
	(0.027120)	(0.048249)	(0.000042)	(0.000013)	(0.000615)
Host * Dist. camp	0.000003***	0.000002**	0.000000	0.000000	0.000000***
	(0.000001)	(0.000001)	(0.000000)	(0.000000)	(0.000000)
Observations	6,907	5,053	12,256	12,256	12,256
Number of districts	85	85	85	85	85
PANEL B: Without units with concentration camps					
Host	-0.145786***	-0.126398*	-0.000047	-0.000001	-0.006686***
	(0.053456)	(0.071081)	(0.000046)	(0.000012)	(0.001594)
Observations	6,837	5,001	12,131	12,131	12,131
Number of districts	85	85	85	85	85
Outcome mean in 1977-2019	0.75	0.75	0.000062	0.000025	0.000552

Note: This table shows the IV estimates of being a *host* municipality (between 1939 and 1975) on social participation from 1977 to 2019 when accounting for the exposure to concentration camps. The construction of reservoirs implied a high demand of prisoners from concentration camps. The data on the location of camps come from *Los colonos de la España Verde de Franco's project*. In panel A, I include an interaction to the distance to the closest concentration camp. In Panel B, I remove from the sample the municipalities with at least one concentration camp. The dataset is at the municipality and year level (for the years of elections for the turnout outcomes). Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.2.5: Reservoir location and size: analysis of determinants

VARIABLES	(1) Reservoir	(2) Reservoir-area	(3) Reservoir	(4) Reservoir-area	(5) Reservoir	(6) Reservoir-area
PANEL A: All municipalities - reservoir in dictatorship						
Rainfall	0.023102***	0.039141	0.023378***	0.066856	0.023370***	0.069361
	(0.002414)	(0.046823)	(0.002410)	(0.046945)	(0.002410)	(0.046901)
Temperature	0.002167	-0.298469***	0.002794	-0.299712***	0.002745	-0.282994***
	(0.002020)	(0.061416)	(0.002019)	(0.060903)	(0.002020)	(0.061492)
Altitude	0.011067***	-0.202269***	0.013052***	-0.187358***	0.013121***	-0.182733***
	(0.002232)	(0.052730)	(0.002249)	(0.052421)	(0.002251)	(0.052410)
Conflict			0.025162***	0.316483***	0.024821***	0.370802***
			(0.003888)	(0.079110)	(0.003922)	(0.084317)
Intensity conflict					0.000007	-0.000469*
					(0.000010)	(0.000254)
Observations	10,738	896	10,738	896	10,738	896
Number of provinces	13	9	13	9	13	9
PANEL B: Municipalities with a planned reservoir in 1933						
Rainfall	0.013775	0.054993***	0.128150***	0.021462*	0.157410***	-0.080273***
	(0.032913)	(0.015641)	(0.036097)	(0.011858)	(0.037286)	(0.004351)
Temperature	-0.094278***	-0.048477***	-0.037606	-0.064095***	-0.025903	-0.027167***
	(0.025603)	(0.011534)	(0.025956)	(0.008589)	(0.026072)	(0.002725)
Altitude	0.017732	0.001653	0.014523	0.067233***	0.003616	0.100379***
	(0.036814)	(0.011766)	(0.035088)	(0.010315)	(0.035007)	(0.003190)
Conflict			0.312829***	0.223357***	0.382698***	0.550040***
			(0.048859)	(0.019106)	(0.054501)	(0.009918)
Intensity conflict					-0.000726***	-0.028423***
					(0.000259)	(0.000706)
Observations	406	168	406	168	406	168
Number of provinces	7	5	7	5	7	5
PANEL C: Municipalities without reservoirs from 1936 to 1975 - reservoirs in democracy						
Rainfall	0.009277**	0.250011***	0.009610**	0.294826***	0.009609**	0.293227***
	(0.003773)	(0.070219)	(0.003762)	(0.069835)	(0.003762)	(0.069740)
Temperature	0.001009	-0.242413***	0.002119	-0.243094***	0.002034	-0.220773***
	(0.003072)	(0.072618)	(0.003066)	(0.071547)	(0.003068)	(0.072585)
Altitude	-0.014930***	-0.174789**	-0.010926***	-0.124370*	-0.010813***	-0.089978
	(0.003418)	(0.072495)	(0.003442)	(0.072224)	(0.003445)	(0.074775)
Conflict			0.048392***	0.497865***	0.047819***	0.591292***
			(0.005938)	(0.105738)	(0.005987)	(0.118436)
Intensity conflict					0.000011	-0.000917*
					(0.000015)	(0.000527)
Observations	10,430	714	10,430	714	10,430	714
Number of provinces	13	9	13	9	13	9

Note: This table shows the OLS estimates of the effect of a set of factors on the location and area or size of a reservoir. I look at average rainfall and temperature (from 1931 to 1932), altitute, having been exposed to conflict, and intensity of conflict. I proxy the violence or conflict perpetrated during the dictatorship (1936-1975) as having or not a mass grave, and conflict intensity as the total number of bodies exhumed from the mass graves in a given municipality m from the Spanish Ministry of Justice. I regress a dummy variable on the geographic-climatic and conflict variables in columns (1) and (3). The Reservoir variable is one if municipality m has a reservoir, 0 otherwise. In columns (2) and (4), the outcome is the reservoir area in squared meter. I include province fixed effects. Panel A shows the effects on the probability of constructing a reservoir during the dictatorship. Panel B repeat the exercise of Panel A in a sample of municipalities with a planned reservoir in 1933. In Panel C shows how the factors understudy affect the probability of having a reservoir during the democracy in a sample of municipalities without a reservoir built by Franco. The dataset is at the municipality and year level (for the years of elections for the turnout outcomes). Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.2.6: Alternative approach: planned reservoirs sample

VARIABLES	(1) Gen. Elec. Turnout	(2) Mun. Elec. Turnout	(3) Cooperatives	(4) Agrarian Coop.	(5) Associations
PANEL A: OLS estimates in a sample of planned reservoirs					
Host	-0.000426	0.009378	-0.000010	0.000000	-0.000560***
	(0.003690)	(0.007057)	(0.000021)	(0.000015)	(0.000195)
Observations	4,570	3,312	8,130	8,130	8,130
Number of districts	63	63	63	63	63
PANEL B: OLS estimates in a sample of planned and constructed reservoirs					
Host	0.008130	0.040768***	-0.000011	0.000014	0.000152
	(0.005927)	(0.011353)	(0.000038)	(0.000021)	(0.000165)
Observations	2,320	1,701	4,122	4,122	4,671
Number of districts	46	46	46	46	50
Outcome mean in 1977-2019	0.75	0.75	0.000062	0.000025	0.000552

Note: This table shows the OLS estimates of being a *host* municipality (between 1939 and 1975) on social participation from 1977 to 2019 in a sample of planned reservoirs. I regress social participation on a dummy variable on the treatment: 1 if municipality m was a *Host* municipality during the Spanish dictatorship (1939-1975), 0 otherwise. In panel A, I include all the bordering municipalities to a planned reservoir, regardless if constructed or not. In Panel B, I restrict my sample to bordering municipalities to planned and constructed reservoirs. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.2.7: Alternative cooperatives and associations outcomes

VARIABLES	(1) Cooperatives	(2) Agrarian Coop.	(3) Associations	(4) Cooperatives	(5) Agrarian Coop.	(6) Associations
Host	-0.001136	-0.000239	-0.281957***	-0.000035	0.000007	-0.008733***
	(0.000935)	(0.000180)	(0.072708)	(0.000040)	(0.000012)	(0.002280)
Observations	12,256	12,256	12,256	12,256	12,256	12,256
Outcome mean in 1977-2019	0.000062	0.000025	0.000552	0.000062	0.000025	0.000552
Number of districts	85	85	85	85	85	85

Note: This table shows the estimates of being a *host* municipality (between 1939 and 1975) on two alternative outcomes for cooperatives and associations. In Columns (1), (2), and (3) I mesure cooperatives using the absolute number of cooperatives and associations in a municipality relative to its population from 1977 to 2018. Columns (4), (5), and (6) show the effects on the number of cooperatives and associations created in a year relative to the native population at the moment of the treatment.. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.2.8: Alternative approach: planned reservoirs sample

VARIABLES	(1) Gen. Elec. Turnout	(2) Mun. Elec. Turnout	(3) Cooperatives	(4) Agrarian Coop.	(5) Associations
Host	-0.132661***	-0.116588**	-0.000047	-0.000000	-0.006232***
	(0.037205)	(0.058037)	(0.000043)	(0.000012)	(0.001048)
Host x $\frac{AdjustedIDP}{native}$	0.028556***	0.024617*	-0.000005	0.000002	0.001094***
	(0.008684)	(0.013589)	(0.000009)	(0.000004)	(0.000319)
Observations	4,570	3,312	8,130	8,130	8,130
Number of districts	63	63	63	63	63
Outcome mean in 1977-2019	0.75	0.75	0.000062	0.000025	0.000552

Note: This table shows the IV estimates of the effect of the intensity of the treatment on being a *host* municipality on social participation by improving the treatment intensity. I adjust my treatment intensity measure by using the total births and deaths per capita. This information is at the province and year level and comes from the Spanish Statistical Office. This data are at the province level. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.2.9: Balance test in pre-treatment characteristics between treated control municipalities

	Mean	Mean	Diff (2) - (1)
	Conterfactual	Treatment	
Population 1900	1,663.147	1,669.612	331.740
	(5,566.975)	(3,082.513)	(220.094)
Population 1940	2,121.803	1,732.879	112.213
	(12,463.813)	(4,745.279)	(286.943)
Women Share (%)	0.494	0.456	-0.029
	(0.072)	(0.109)	(0.022)
Illiterate Share (%)	0.710	0.669	-0.036
	(0.124)	(0.175)	(0.035)
Married Share (%)	0.365	0.335	-0.006
	(0.072)	(0.092)	(0.019)
Widowed Share (%)	0.079	0.075	-0.002
	(0.022)	(0.023)	(0.005)
Widow Share (%)	0.108	0.108	0.004
	(0.031)	(0.026)	(0.006)
New cooperatives in 1945 per capita	0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)
New agricultural cooperatives in 1945 per capita	0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)
New associations in 1945 per capita	0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)
Observations	16,631	5,125	21,756

Notes: This table reports descriptive statistics for the main variables and sample considered in the analysis. The analysis covers 1,695 municipalities (for the 14 general elections from 1977 to 2019). Pre-treatment characteristics are from the 1940 Population census. Municipality characteristics are from 1991 and 2011 Spanish Population Census.

Table A.2.10: Exclusion restriction tests

VARIABLES	(1) Gen. Elec. Turnout	(2) Mun. Elec. Turnout	(3) Cooperatives	(4) Agrarian Coop.	(5) Associations
PANEL A: Controlling for modern-day rainfall					
Host	-0.135358***	-0.119459*	-0.000049	0.000016	-0.006311***
	(0.048232)	(0.068660)	(0.000048)	(0.000015)	(0.001550)
Observations	6,907	5,053	12,256	12,256	12,256
Number of districts	85	85	85	85	85
PANEL B: OLS estimates in the counterfactual sample					
$Plan1933_{mdp} * \frac{Area_{mdp}}{Distance_{mdp}}$	0.001388	0.002426	0.000002	-0.000001	-0.000009
	(0.001143)	(0.002341)	(0.000012)	(0.000005)	(0.000047)
Observations	5,355	3,912	9,501	9,501	9,501
Number of districts	76	76	76	76	76
Outcome mean in 1977-2019	0.75	0.75	0.000062	0.000025	0.000552

Note: This table shows two validation test of the exclusion restriction. I measure social participation from 1977 to 2019 as "voter turnout in general elections" (column (1)), "voter turnout in municipal elections" (column (2)), "number of cooperatives created yearly" (column (3)), "number of cooperatives in the agricultural created yearly" (column (4)), and "number of associations created yearly" (column (5)). I regress social participation on a dummy variable on the treatment: 1 if municipality m was a *Host* municipality during the Spanish dictatorship (1939-1975), 0 otherwise. I control for pre-treatment municipality characteristics (share female, share illiterate, and share single in 1940), district fixed effects and province-year interactions fixed effects. Standard errors are clustered at the district level. In panel A, I control for modern-day rainfall. Panel B shows the OLS estimates of a regression of the instrument on social participation outcomes. I use an interaction between the margin of whether the closest reservoir to municipality m was planned or not in the 1933 Plan to the area of the closest reservoir weighted by the inverse of the distance to the closest reservoir as my instrument. The dataset is at the municipality and year level (for the years of elections for the turnout outcomes). Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.2.11: Heterogeneity analysis on social participation

VARIABLES	(1) Gen. Elec. Turnout	(2) Mun. Elec. Turnout	(3) Cooperatives	(4) Agrarian Coop.	(5) Associations
PANEL A: Heterogenity by *survivors* share					
Host	-0.312520**	-0.233481*	-0.000140	-0.000031	-0.026675***
	(0.146166)	(0.132820)	(0.000180)	(0.000057)	(0.005015)
Host * *survivors* share	0.525675**	0.406086*	0.000256	0.000084	0.054481***
	(0.256992)	(0.235893)	(0.000383)	(0.000127)	(0.011912)
Observations	6,907	5,053	12,256	12,256	12,256
PANEL B: Heterogenity by decade of treatment					
Host	-0.146086***	-0.125955*	-0.000046	-0.000000	-0.006705***
	(0.053703)	(0.071344)	(0.000045)	(0.000012)	(0.001598)
Host	-0.139381***	-0.121829*	-0.000050	-0.000003	-0.006353***
	(0.044887)	(0.065875)	(0.000045)	(0.000011)	(0.001210)
Host * 1930s	0.142034***	0.136649**	-0.000030	-0.000024**	0.006428***
	(0.044945)	(0.067250)	(0.000034)	(0.000010)	(0.001187)
Host * 1940s	0.042133	-0.020637	0.000197**	0.000144	0.003975**
	(0.047754)	(0.050308)	(0.000090)	(0.000089)	(0.001566)
Host * 1960s	0.091413**	0.057608	-0.000026	-0.000028	0.004870***
	(0.041660)	(0.054065)	(0.000043)	(0.000023)	(0.001255)
Observations	6,907	5,053	12,256	12,256	12,256
PANEL C: Heterogenity by decade of outcomes					
Host	-0.554032***	-0.281207**	-0.000160	-0.000065	-0.028215***
	(0.209448)	(0.133879)	(0.000192)	(0.000057)	(0.008024)
Host * 1980s	0.457998**		0.000116	0.000069	0.024334***
	(0.183565)		(0.000161)	(0.000056)	(0.007561)
Host * 1990s	0.467517**	0.221881**	0.000133	0.000072	0.024528***
	(0.183176)	(0.112180)	(0.000184)	(0.000057)	(0.007346)
Host * 2000s	0.480679***	0.221821**	0.000209	0.000131	0.024630***
	(0.183917)	(0.109656)	(0.000221)	(0.000127)	(0.007426)
Host * 2010s	0.490049***	0.246751**	0.000133	0.000068	0.024561***
	(0.186460)	(0.110736)	(0.000178)	(0.000056)	(0.007350)
Observations	6,907	5,053	12,256	12,256	12,256
Outcome mean in 1977-2019	0.75	0.75	0.000062	0.000025	0.000552
Number of districts	85	85	85	85	85

Note: This table shows the IV estimates of being a *host* municipality (between 1939 and 1975) on social participation from 1977 to 2019 with three heterogeneity analysis. I measure social participation as "voter turnout in general elections" (column (1)), "voter turnout in municipal elections" (column (2)), "number of cooperatives created yearly" (column (3)), "number of cooperatives in the agricultural created yearly" (column (4)), and "number of associations created yearly" (column (5)). I regress social participation on a dummy variable on the treatment: 1 if municipality m was a *Host* municipality during the Spanish dictatorship (1939-1975), 0 otherwise. I control for pre-treatment municipality characteristics (share female, share illiterate, and share single in 1940), district fixed effects and province-year interactions fixed effects. Standard errors are clustered at the district level. The instrument is an interaction between the margin of whether the closest reservoir to municipality m was planned or not in the 1933 Plan to the area of the closest reservoir weighted by the inverse of the distance to the closest reservoir. In panel A, I interact the treatment to the total population share born before the arrival of population that still live at the *host* municipalities. I call them *survivors*, either because they were the new comers or the native. In Panel B, I interact the treatment to the timing of the treatment by decade (1936-1962). Finally, Panel C produces the heterogeneous results by decade in the outcome (1977-2019). The dataset is at the municipality and year level (for the years of elections for the turnout outcomes). Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.2.12: Timing of treatment heterogeneity

	(1)	(2)	(3)	(4)	(5)
VARIABLES	Gen. Elec. Turnout	Mun. Elec. Turnout	Cooperatives	Agrarian Coop.	Associations
PANEL A: Baseline sample					
Host	-0.146086***	-0.125955*	-0.000046	-0.000000	-0.006705***
	(0.053703)	(0.071344)	(0.000045)	(0.000012)	(0.001598)
Observations	6,907	5,053	12,256	12,256	12,256
Number of districts	85	85	85	85	85
PANEL B: Dropping treated in 1930s					
Host	-0.137028***	-0.120861*	-0.000045	-0.000001	-0.006798***
	(0.045090)	(0.067636)	(0.000044)	(0.000011)	(0.001753)
Observations	6,725	4,914	11,933	11,933	11,933
Number of districts	84	84	84	84	84
PANEL C: Dropping treated in 1940s					
Host	-0.137115***	-0.120614*	-0.000046	-0.000002	-0.006857***
	(0.045608)	(0.068130)	(0.000045)	(0.000011)	(0.001770)
Observations	6,543	4,775	11,614	11,614	11,614
Number of districts	80	80	80	80	80
PANEL D: Dropping treated in 1950s					
Host	0.064337	0.101881	0.000194	-0.000052	-0.001370
	(0.114515)	(0.179481)	(0.000279)	(0.000085)	(0.002831)
Observations	5,718	4,185	10,145	10,145	10,145
Number of districts	79	79	79	79	79
Outcome mean in 1977-2019	0.75	0.75	0.000062	0.000025	0.000552

Note: This table shows the effect of being a *host* municipality (between 1939 and 1975) on social participation from 1977 to 2019 by dropping the treated municipalities by decade of treatment. I repeat the analysis of equation (1). Panel A shows the baseline results. In panel B, I drop the municipalities treated in the 1930s. Panel C documents the IV estimates when I drop the municipalities treated in the 1930s and 1940s. Finally, panel D drops the treated municipalities in 1950s. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.2.13: Falsification tests

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Cooperatives	Agrarian Coop.	Associations	Cooperatives	Agrarian Coop.	Associations
PANEL A: Effect on the years preceding the treatment						
Host	-0.000006	-0.000103	-0.000009	0.000084	-0.000191	-0.002728
	(0.000012)	(0.000290)	(0.000008)	(0.000121)	(0.000163)	(0.003363)
Observations	21,998	21,998	21,998	21,998	21,998	21,998
Number of districts	85	85	85	85	85	85
PANEL B: Host democracy on social participation in 1977						
VARIABLES	Gen. Elec. Turnout	Mun. Elec. Turnout	Cooperatives	Agrarian Coop.	Associations	
host in democracy	0.677296	1.710512	0.000021	-0.000514	0.000911	
	(0.768975)	(2.248787)	(0.000033)	(0.001425)	(0.000728)	
Observations	385	276	416	416	416	
Number of districts	76	68	77	77	77	
PANEL C: Placebo counterfactual						
Host	-0.091573***	-0.118687**	0.000019	0.000020**	-0.004030***	
	(0.025284)	(0.053119)	(0.000014)	(0.000010)	(0.000718)	
Observations	3,508	2,538	6,241	6,241	6,241	
Number of district	54	54	54	54	54	
Outcome mean in 1977-2019	0.75	0.75	0.000062	0.000025	0.000552	

Note: This table shows the effect of being a *host* municipality (between 1939 and 1975) on social participation from 1977 to 2019 by dropping the treated municipalities by decade of treatment. I repeat the analysis of equation (1). In panel A, I look at the impact on cooperatives and associations in the year strictly preceding the treatment. In Panel B, I repeat the analysis of equation (1) with reservoirs constructed after 1977 and look at the potential effects of outcomes in 1977. Finally, in panel C I compare treated municipalities against bordering municipalities to a municipality where a reservoir was projected but never built (placebo counterfactual). Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.2.14: Potential confounding effects of violence

VARIABLES	(1) Gen. Elec. Turnout	(2) Mun. Elec. Turnout	(3) Cooperatives	(4) Agrarian Coop.	(5) Associations
PANEL A: OLS estimates intensity of violence on social participation					
Violence intensity	-0.000041*	-0.000135***	-0.000000	-0.000000	-0.000001
	(0.000021)	(0.000039)	(0.000000)	(0.000000)	(0.000001)
Observations	6,907	5,053	12,256	12,256	12,256
Number of districts	85	85	85	85	85
PANEL B: Controlling for the intensity of violence					
Host	-0.145937***	-0.124886*	-0.000047	0.000000	-0.006695***
	(0.053926)	(0.071481)	(0.000046)	(0.000012)	(0.001602)
Observations	6,907	5,053	12,256	12,256	12,256
Number of districts	85	85	85	85	85
PANEL C: Without units affected by violence					
Host	-0.130943***	-0.117592*	-0.000063	-0.000006	-0.006207***
	(0.041193)	(0.065401)	(0.000054)	(0.000014)	(0.001048)
Observations	5,004	3,619	8,878	8,878	8,878
Number of counties	80	80	80	80	80
Outcome mean in 1977-2019	0.75	0.76	0.86	0.49	

Note: This table shows the IV estimates of being a *host* municipality (between 1939 and 1975) on social participation from 1977 to 2019 when accounting for intensity of violence during the dictatorship. I measure social participation as "voter turnout in general elections" (column (1)), "voter turnout in municipal elections" (column (2)), "number of cooperatives created yearly" (column (3)), "number of cooperatives in the agricultural created yearly" (column (4)), and "number of associations created yearly" (column (5)). I regress social participation on a dummy variable on the treatment: 1 if municipality m was a *Host* municipality during the Spanish dictatorship (1939-1975), 0 otherwise. I control for pre-treatment municipality characteristics (share female, share illiterate, and share single in 1940), district fixed effects and province-year interactions fixed effects. Standard errors are clustered at the district level. The instrument is an interaction between the margin of whether the closest reservoir to municipality m was planned or not in the 1933 Plan to the area of the closest reservoir weighted by the inverse of the distance to the closest reservoir. Panel A shows the OLS estiamtes of the effect of violence intensity on social participation. In panel B, I include level of violence fixed effects. I measure the violence perpetrated during the dictatorship (1936-1975) as total number of bodies exhumed from the mass graves in a given municipality m from the Spanish Ministry of Justice. In Panel C, I remove from the sample the municipalities with at least one mass grave. The dataset is at the municipality and year level (for the years of elections for the turnout outcomes). Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.2.15: Economic development test

VARIABLES	(1) Gen. Elec. Turnout	(2) Mun. Elec. Turnout	(3) Cooperatives	(4) Agrarian Coop.	(5) Associations
PANEL A: Dropping poorest region - Castilla-La Mancha					
Host	-0.146086***	-0.125955*	-0.000046	-0.000000	-0.006705***
	(0.053703)	(0.071344)	(0.000045)	(0.000012)	(0.001598)
Observations	6,907	5,053	12,256	12,256	12,256
Number of districts	85	85	85	85	85
PANEL B: Dropping second poorest region - Castellón					
Host	-0.143822***	-0.125133*	-0.000047	-0.000000	-0.006685***
	(0.051027)	(0.070366)	(0.000046)	(0.000012)	(0.001573)
Observations	6,879	5,033	12,208	12,208	12,208
Number of districts	84	84	84	84	84
PANEL C: Dropping third poorest region - Cantabria					
Host	-0.136944***	-0.121056*	-0.000044	0.000002	-0.006792***
	(0.043409)	(0.066180)	(0.000045)	(0.000013)	(0.001674)
Observations	6,823	4,989	12,110	12,110	12,110
Number of districts	83	83	83	83	83
PANEL D: Dropping forth poorest region - Castilla y León					
Host	-0.121689***	-0.094437***	-0.000044	0.000003	-0.006217***
	(0.023409)	(0.021411)	(0.000048)	(0.000013)	(0.000610)
Observations	6,403	4,690	11,366	11,366	11,366
Number of districts	74	74	74	74	74
PANEL E: Dropping fifth poorest region - La Rioja					
Host	0.028676	0.061938	0.000111	-0.000066	-0.001919*
	(0.064627)	(0.093957)	(0.000154)	(0.000078)	(0.001039)
Observations	5,885	4,321	10,442	10,442	10,442
Number of districts	65	65	65	65	65
PANEL F: Dropping sixth poorest region - Aragón					
Host	-0.137337	-0.220272	-0.000401	-0.000223	-0.000452
	(0.309277)	(0.201337)	(0.000468)	(0.000281)	(0.001820)
Observations	2,613	1,928	4,605	4,605	4,605
Number of districts	24	24	24	24	24
Outcome mean in 1977-2019	0.75	0.75	0.000062	0.000025	0.000552

Note: This table shows the effect of being a *host* municipality (between 1939 and 1975) on social participation from 1977 to 2019 by removing region by region in order of lowest development. I measure development using GDP per capita at regional level for 2018 from the Spanish Statistical Office. I repeat the analysis of equation (1). Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.2.16: Potential confounding effects of economic development

VARIABLES	(1) Gen. Elec. Turnout	(2) Mun. Elec. Turnout	(3) Cooperatives	(4) Agrarian Coop.	(5) Associations
PANEL A: Controlling for unemployment population share					
Host	-0.114229***	-0.058610	-0.000028	0.000007	-0.007077***
	(0.030852)	(0.044291)	(0.000039)	(0.000015)	(0.001691)
Observations	5,848	4,414	11,958	11,958	11,958
Number of districts	84	84	84	84	84
Number of districts	85	85	85	85	85
PANEL B: Controlling for agricultural employment population share					
Host	-0.112851***	-0.056720	-0.000042	0.000000	-0.007155***
	(0.031508)	(0.045340)	(0.000043)	(0.000012)	(0.001720)
Observations	5,848	4,414	11,958	11,958	11,958
Number of districts	84	84	84	84	84
height PANEL C: Controlling for depopulation in 1950s					
Host	-0.144157***	-0.126856*	-0.000048	-0.000003	-0.006710***
	(0.050425)	(0.072162)	(0.000046)	(0.000012)	(0.001587)
Observations	6,907	5,053	12,256	12,256	12,256
Number of districts	85	85	85	85	85
Outcome mean in 1977-2019	0.75	0.76	0.86	0.49	

Note: This table shows the IV estimates of being a *host* municipality (between 1939 and 1975) on social participation from 1977 to 2019 when controlling for local labour market characteristics. I repeat equation (1). In panel A, I measure local labour market with data on share of unemployed population. In Panel B, I control for population share employed in agricultural sector. I control for depopulation in the 50s in Panel C. There was a significant out-migration flow during the 50s in Spain from backward regions to the leading areas, with rural families heading for urban areas (Pinilla and Sáez 2016) I use population changes between the population census in 1950 and 1960 to try to capture this variation. The dataset is at the municipality and year level (for the years of elections for the turnout outcomes). Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.2.17: Alternative treatment definition: *host* municipalities in democracy (1976-2019)

VARIABLES	(1) Gen. Elec. Turnout	(2) Mun. Elec. Turnout	(3) Cooperatives	(4) Agrarian Coop.	(5) Associations
Host in democracy	0.075247	-0.024477	0.000418	0.000247	-0.001873
	(0.167744)	(0.162204)	(0.000407)	(0.000261)	(0.002620)
Outcome mean in 1977-2019	0.75	0.76	0.86	0.49	
Observations	5,775	4,196	10,252	10,252	10,252
Number of districts	77	77	77	77	77

Note: This table shows the estimates of being a *host* municipality (between 1976 and 2018) on social participation from 1977 to 2019. I repeat equation (1). The treatment now only includes the bordering municipalities to four reservoirs destroying a municipality during the democracy. Only 61 municipalities are defined as *host municipalities*. Control municipalities are the bordering municipalities to a reservoir not destroying a municipality. I include all the reservoirs built before the outcome year. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.2.18: Sample test: same timing across outcomes

VARIABLES	(1) Gen. Elec. Turnout	(2) Mun. Elec. Turnout	(3) Cooperatives	(4) Agrarian Coop.	(5) Associations
PANEL A: General elections timing constant					
Host	-0.146086***	-0.222088*	-0.000001	0.000009	-0.005495***
	(0.053703)	(0.121486)	(0.000018)	(0.000023)	(0.001165)
Observations	6,907	1,806	4,352	4,352	4,352
Number of districts	85	85	85	85	85
PANEL B: Municipal elections timing constant					
Host	-0.205520**	-0.125955*	-0.000071	-0.000025	-0.004330***
	(0.083222)	(0.071344)	(0.000090)	(0.000021)	(0.000772)
Observations	1,976	5,053	3,410	3,410	3,410
Number of districts	85	85	85	85	85
Outcome mean in 1977-2019	0.75	0.76	0.86	0.49	

Note: This table shows the estimates of being a *host* municipality (between 1976 and 2018) on social participation from 1977 to 2019 keeping constant the timing across columns. I repeat equation (1). I Panel A I use the years of the general elections. In Panel B I keep constant the timing using the years of the municipal elections. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.2.19: Alternative treatment definition: *extended host* and municipalities

VARIABLES	(1) Gen. Elec. Turnout	(2) Mun. Elec. Turnout	(3) Cooperatives	(4) Agrarian Coop.	(5) Associations
PANEL A: Treated= *host* and *extended host* mun.					
Host and *extended host*	-0.232961**	-0.146399	-0.000073	-0.000007	-0.007790**
	(0.114433)	(0.153009)	(0.000067)	(0.000023)	(0.003195)
Observations	8,587	6,274	15,137	15,137	15,137
Number of districts	89	89	88	88	88
PANEL B: Treated= *extended host* mun.					
Extended host	0.134723	0.867406	0.001525	-0.001373	0.042613
	(2.395695)	(3.188446)	(0.005852)	(0.004652)	(0.141788)
Observations	7,035	5,133	12,382	12,382	12,382
Number of districts	86	86	85	85	85
PANEL C: Treated= *host* and destroyed mun.					
Host and destroyed	0.081845	-0.258223	0.000097	-0.000018	-0.002055**
	(0.177895)	(0.258705)	(0.000117)	(0.000031)	(0.000929)
Observations	7,341	5,383	13,031	13,031	13,031
Number of districts	85	85	85	85	85
Outcome mean in 1977-2019	0.75	0.76	0.86	0.49	

Note: This table shows the estimates for alternative definitions of the treatment (between 1976 and 2018) on social participation from 1977 to 2019. I repeat equation (1). In Panel A, I extend the treatment to the surrounding municipalities to the treated or *host* municipalities. Panel B shows the results restricting the treatment definition to the *extended host* municipalities. I call them *extended host* municipalities. In Panel C, I include the *host* and destroyed municipalities in the sample Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.2.20: Alternative specification

VARIABLES	(1) Gen. Elec. Turnout	(2) Mun. Elec. Turnout	(3) Cooperatives	(4) Agrarian Coop.	(5) Associations
PANEL A: Clusters at province level					
Host	-0.115630***	-0.132791**	-0.000209**	-0.000019	-0.001228*
	(0.044144)	(0.066406)	(0.000090)	(0.000024)	(0.000736)
Number of provinces	14	14	14	14	14
PANEL B: Clusters at region level					
Host	-0.111109***	-0.126582**	-0.000198**	-0.000019	-0.001290
	(0.035374)	(0.055143)	(0.000083)	(0.000026)	(0.000876)
Number of regions	8	8	8	8	8
PANEL C: Province and Year-Region fixed effects					
Host	-0.115630***	-0.132791**	-0.000209**	-0.000019	-0.001228*
	(0.044144)	(0.066406)	(0.000090)	(0.000024)	(0.000736)
Number of provinces	14	14	14	14	14
PANEL D: Region and Year fixed effects					
Host	-0.111109***	-0.126582**	-0.000198**	-0.000019	-0.001290
	(0.035374)	(0.055143)	(0.000083)	(0.000026)	(0.000876)
Number of regions	8	8	8	8	8
PANEL E: Modern-day covariates					
Host	-0.152857***	-0.186791**	-0.000002	0.000015	-0.000681**
	(0.058052)	(0.079322)	(0.000049)	(0.000028)	(0.000329)
Observations	6,907	5,053	12,256	12,256	12,256
Outcome mean in 1977-2019	0.75	0.76	0.86	0.49	

Note: This table shows the estimates for alternative specification to equation (1). In Panel A and B, I cluster the error terms at the province or regional level, respectively. Panel C shows the results by including province and year-region fixed effects. In Panel D, I control for region and time fixed effects. In Panel E, I control for post-treatment socio-demographic characteristics (older than 64, women, educated, foreign population shares) from the population census. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.2.21: Effects on satisfaction with Franco's government: Falanguist vote share

VARIABLES	(1) Falanguist Gen. Elec.	(2) Falang. G.E.	(3) Falang. G.E.	(4) Falang. Mun. Elec.	(5) Falang. M.E.	(6) Falang. M.E.
PANEL A: OLS estimates						
Host	-0.000060	-0.000051	-0.000078	0.000005	0.000010	0.000001
	(0.000099)	(0.000116)	(0.000124)	(0.000021)	(0.000025)	(0.000027)
Host x $\frac{IDP}{native}$		-0.000006			-0.000004	
		(0.000039)			(0.000008)	
PANEL B: IV estimates						
Host	0.000978***	0.000910***	0.002152***	0.000038	0.000035	0.000061
	(0.000264)	(0.000217)	(0.000788)	(0.000050)	(0.000045)	(0.000083)
inflow_vs_native		-0.000172***			-0.000008	
		(0.000058)			(0.000009)	
Observations	6,907	6,907	6,907	5,053	5,053	5,053
Outcome mean in 1977-2019	0.00039	0.00039	0.00039	0.00003	0.00003	0.00003
Number of districts	85	85	85	85	85	85

Note: This table shows the estimates of being a *host* municipality (between 1976 and 2018) on satisfaction with Franco's government from 1977 to 2019. I repeat equation (1). I use data on voting share for Falangist parties from 1977 to 2019 as a proxy of support or satisfaction with Franco's government. Falangist parties are far-right parties align with Franco's ideology. Falangism (Spanish: falangismo) was the political ideology of two political parties in Spain that were known as the *Falange*, namely first the *Falange Española de las Juntas de Ofensiva Nacional Sindicalista* (FE de las JONS) and afterwards the *Falange Española Tradicionalista y de las Juntas de Ofensiva Nacional Sindicalista* (FET y de las JONS). Columns (1), (2) and (3) shows the results on the general elections. Columns (4), (5) and (6) produce the results for the municipal elections. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.2.22: Effect of being a *host* municipality in democracy (1976-2019) on trust

VARIABLES	(1) mistrust institution	(2) general trust	(3) mistrust institution	(4) general trust
PANEL A: OLS and IV estimates				
Host in democracy	-0.049459***	-3.937248***	-0.745626***	-3.460112***
	(0.008153)	(0.395735)	(0.253096)	(0.629196)
PANEL B: With treatment intensity				
Host in democracy	-0.000012	0.001849	-0.050576***	-3.141422***
	(0.000011)	(0.000539)	(0.008246)	(0.431271)
Host in demo. x $\frac{IDP}{native}$	-0.000011	0.001230	-0.000012	0.001849***
	(0.011657)	(0.069884)	(0.000013)	(0.000439)
Observations	1,120	476	1,008	434
Number of provinces	9	8	9	7

Note: This table produces the effect of being a *host* municipality (between 1976 and 2019) on trust from 1989 to 2015. I measure trust as "institutional mistrust" (columns (1) and (3)), and "general trust" (column (2) and (4)). I proxy institutional trust with the average share of individuals who reported mistrust in the national government for each municipality, and general trust with the average general trust. Trust data come from the Spanish Sociological Research Center (CIS, Centro de Investigaciones Sociológicas). I regress trust on a dummy variable on the treatment: 1 if municipality m was a *host* municipality during the Spanish dictatorship (1939-1975), 0 otherwise. I control for pre-treatment municipality characteristics (share female, share illiterate, and share single in 1940), district fixed effects and province-year interactions fixed effects. Standard errors are clustered at the district level. In panel A, I produce the OLS (columns (1) and (2)) and IV (columns (3) and (4)) estimates. In Panel B, I interact the treatment to the treatment intensity (number of forcibly displaced population affected by a reservoir relative to the native population to measure treatment intensity). The dataset is at the municipality and year level (for the years of elections for the turnout outcomes). Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.2.23: Correlations trust on social participation (1977-2019)

VARIABLES	(1) Gen. Elec. Turnout	(2) Mun. Elec. Turnout	(3) Cooperatives	(4) Agrarian Coop.	(5) Associations
PANEL A: Institutional mistrust					
Inst. mistrust	0.059891***	0.096044**	-0.000015	-0.000014	0.001029***
	(0.020236)	(0.044216)	(0.000126)	(0.000114)	(0.000194)
Inst. mistrust * Host	0.164479***	-0.105933	0.000038	-0.000001	-0.001601***
	(0.060131)	(0.122857)	(0.000378)	(0.000342)	(0.000582)
PANEL B: Institutional mistrust - treatment intensity					
Inst. mistrust	0.057903***	0.092643**	-0.000015	-0.000014	0.001072***
	(0.019910)	(0.043964)	(0.000121)	(0.000109)	(0.000191)
Inst. mistrust * Host	1.084020***	0.920392***	0.000326	0.000118	-0.002523
	(0.162678)	(0.333942)	(0.000994)	(0.000898)	(0.001566)
Inst. mistrust * Host * $\frac{IDP}{native}$	-2.676174***	-2.983866***	-0.000839	-0.000344	0.002638
	(0.441034)	(0.903628)	(0.002679)	(0.002421)	(0.004222)
PANEL C: Institutional mistrust - survivors share					
Inst. mistrust	0.059730***	0.095174**	-0.000015	-0.000014	0.001071***
	(0.020196)	(0.044032)	(0.000121)	(0.000109)	(0.000191)
Inst. mistrust * Host	0.276792***	0.146245	0.000113	0.000093	-0.002420**
	(0.077354)	(0.151547)	(0.000629)	(0.000569)	(0.000992)
Inst. mistrust * Host * *survivors* share	-1.113877**	-2.682198***	-0.000352	-0.000436	0.003748
	(0.484070)	(0.951279)	(0.002384)	(0.002153)	(0.003756)
Observations	1,287	996	2,372	2,372	2,372
Number of districts	47	47	47	47	47
PANEL D: General trust					
Gen. trust	0.008733***	0.002626	0.000007	0.000007*	0.000094***
	(0.002003)	(0.003200)	(0.000006)	(0.000004)	(0.000020)
Gen. trust * Host	-0.014292***	-0.029303***	-0.000008	-0.000006	-0.000152***
	(0.004465)	(0.007120)	(0.000014)	(0.000008)	(0.000045)
PANEL E: General trust - treatment intensity					
Gen. trust	0.007219***	0.005336	0.000007	0.000008**	0.000086***
	(0.002113)	(0.003351)	(0.000007)	(0.000004)	(0.000021)
Gen. trust * Host	0.022438	-0.096357***	-0.000015	-0.000028	0.000053
	(0.017613)	(0.027794)	(0.000056)	(0.000032)	(0.000176)
Gen. trust * Host * $\frac{IDP}{native}$	-0.085413**	0.156047**	0.000017	0.000050	-0.000475
	(0.039631)	(0.062562)	(0.000126)	(0.000071)	(0.000395)
PANEL F: General trust - survivors share					
Gen. trust	0.008712***	0.002617	0.000006	0.000007*	0.000094***
	(0.001998)	(0.003204)	(0.000006)	(0.000004)	(0.000020)
Gen. trust * Host	-0.010634**	-0.027857***	-0.000002	-0.000006	-0.000169**
	(0.004946)	(0.007774)	(0.000021)	(0.000012)	(0.000065)
Gen. trust * Host * *survivors* share	-0.026576*	-0.011237	-0.000024	-0.000001	0.000067
	(0.015628)	(0.024082)	(0.000059)	(0.000033)	(0.000185)
Observations	545	424	1,003	1,003	1,003
Number of districts	27	27	27	27	27
Outcome mean in 1977-2019	0.75	0.76	0.86	0.49	

Note: This table shows the OLS estimates of the effect of trust on social participation from 1977 to 2019. I repeat equation (1). I proxy institutional trust with the average share of individuals who reported mistrust in the national government for each municipality, and general trust with the average general trust. Trust data come from the Spanish Sociological Research Center (CIS, Centro de Investigaciones Sociológicas). In Panel A and D, I show the baseline results with an interaction if *host* municipality. In Panel B and E, I look at the treatment intensity, measured as the number of forcibly displaced population affected by a reservoir relative to the native population to measure treatment intensity. In Panel C and F, I add an interaction to the total population share born before the arrival of population that still live at the *host* municipalities. I call them *survivors*, either because they were the new comers or the native. Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Chapter 3

War on Polio Eradication: Reaching the Hard-to-Reach

3.1 Introduction

Disease eradication is the mother of all global health efforts, as everyone worldwide can enjoy the well-being benefits of eradication. Smallpox is the only human disease to be eradicated worldwide. And we are closer than ever to ending polio, but outbreaks and challenges persist (UNICEF 2019).

The increasing number of internally displaced people (IDP) poses new challenges to eradicating polio worldwide. New conflicts are emerging in polio-affected countries, and with it, the number of people forced to flee. The forcibly displaced population climbed to 89.3 million by 2021. This figure is more than double the 42.7 million people who remained forcibly displaced in 2012. Internal displacement in 2021 was markedly higher than in recent years, corresponding to 66% of the total. This is the largest displaced population, but it is also one of the most vulnerable. The vast majority do not receive the medical care they need because they live in regions where the healthcare system has collapsed, or national governments do not protect their health rights. Therefore children in these fragile areas are at higher risk of contracting and spreading polio. (UNHCR 2022).

From a policy perspective, improving our knowledge of the role of IDPs in the incidence of polio is crucial for a better design and efficacy of preventive public policies. A polio-free world will save the global economy USD 45 billion in health costs within the next 20 years (UNICEF 2019). Additionally, eradicating polio could be the prelude to other diseases' eradication.

In this paper, I study the impacts of IDP inflows on polio incidence in host communities and evaluate some of the main potential channels. To tackle this question, I use the mass displacement of the 57% of the population from the conflict-affected

Federally Administered Tribal Areas (F.A.T.A.) to other districts in Pakistan from 2008 to 2022. In a difference-in-differences approach, I compare the polio cases between host and non-host districts before and after 2007. Due to cultural and linguistic barriers, most IDPs settled within the historical region of *Pashtunistan*. I exploit the spatial distribution of districts with respect to the pre-colonial region of *Pashtunistan*'s border to define the host and non-host districts.

I find that districts that received the IDP population have experienced an increase in the probability of polio incidence compared to non-host districts. An IDP inflow increased the likelihood of at least one polio case by 4.1 percentage points (pp) and 0.007 additional cases per 100,000 inhabitants in the host districts. The estimates are statistically significant at the five per cent level. Although negligible, my findings represent an increase of 40% over the mean incidence.

I also look at the intensity of the inflow. To do so, I rely on districts closer to the F.A.T.A. border receiving more IDPs when the total yearly inflows of IDPs increase. More formally, as data on migration flows are unavailable at the district level, I construct a yearly district measure of predicted IDP inflows based on the interaction of the inverse distance to the F.A.T.A. border and the total yearly migration flows from F.A.T.A. to other regions in Pakistan. I show that an increase of one standard deviation in predicted inflows results in 0.001 additional polio cases per 100,000 inhabitants, corresponding to 20% of the mean incidence.

Why are the main effects meaningful? Polio, which only infects humans, has been eliminated in 193 countries. With the transmission of wild-type polio limited to Afghanistan and Pakistan, an official eradication declaration is in sight. In 2005, 28 cases were reported in Pakistan, compared to the 1,147 cases in 1997. Moreover, most host districts had zero or close to zero polio cases before 2007. Ultimately, the estimates I present in this paper capture the impacts of IDP inflows on keeping host communities away from eradicating polio rather than the effects on an extensive increase in polio cases.

I propose three potential mechanisms by which IDP inflow could slow down polio eradication in host communities: a sudden increase in the population in communities with low vaccination rates, the precarious health conditions in host communities, and the congestion of health services in host communities. I use individual-level data from the Demographic and Health Survey (D.H.S.) from 1990 to 2017 to generate supporting evidence. First, I show no statistical differences in vaccination rates between host and non-host districts before and after 2007, with no differences between IDPs and native children. But, I observe a national immunisation rate decrease around 2007 which could increase the susceptibility to transmitting polio in overpopulated communities. Second, IDPs settle in communities with poorer conditions than non-host districts even before the large IDP inflow (i.e. higher number of members and children under five in the household, more likely to live in urban settings, and less likely to have a head of household working). Finally, I observe increased individual demand for health services in host districts after 2007. I measure the individual demand with the share of children with prenatal assistance.

To support the validity of my results, I estimate a dynamic difference-in-differences specification in which I calculate the difference in the polio cases between host and non-host districts on a yearly basis. The exercise supports the validity of the parallel trend assumption and shows similar impacts of IDP inflows to the ones identified in the aggregate regression. I also rule out the existence of confounders' effects from the conflict. For this purpose, I show that the conflict did not indirectly impact host communities' polio cases before the onset of the IDP crises in 2008, which validates my main findings. Additionally, I validate my treatment and counterfactual with alternative definitions. Finally, I test whether reverse causality between IDPs and polio exists.

The situation in Pakistan is not unique. In 2022, polio cases emerged in Malawi and Mozambique– two countries free of the virus for decades. Both countries are the scenario of a forced population movement from the conflicts ongoing in Northern Mozambique and the Eastern Democratic Republic of Congo, respectively. Protected children from diseases are far more likely to have the opportunity to thrive, the chance to learn and the ability to live healthy lives (UNICEF 2023). Therefore, three critical policy implications emerge from this paper. First, millions of forcibly displaced children migrate to camps or host communities. Since families in these settings are often transient, monitoring vaccination rates among these communities make it much harder to reach children with the necessary vaccines. Reaching the hard-to-reach- such as children from mobile and forced migrant populations or in conflict zones- should be a public priority (CDC 2021; UNICEF 2023). Second, poor communities are the host communities of most of the IDPs. An effort to better integrate the IDP population into the health services and labour market should be made to improve the conditions in which they live. Finally, the inflow of new population comes with increased demand for health services. Even if the increase in the demand is modest, in locations where the health delivery or capacity is weak, it can congest the local health services. It is essential to reinforce host communities' health workforce and infrastructure, so locals and newcomers can access health services equally.

This paper adds to the literature that studies the consequences of forcibly displaced populations in host communities (S. Becker and Ferrara 2019) by evaluating the impact on polio incidence, a disease close to being eradicated. As far as I know, the literature on the impacts of hosting displaced people on the spread of diseases is limited (Montalvo and Reynal-Querol 2007; Baez 2011; Ibáñez, Rozo, and Urbina

2021; Ibáñez, Moya, et al. 2023). None of the existing literature examines the effect of hosting internally displaced populations. This population does not cross an international border, making monitoring their health and vaccination status much harder. So, the impact of these inflows may differ from refugee inflows. Understanding the role of IDPs in polio-endemic countries is vital for formulating quick interventions in transit zones, camps and host communities. Additionally, I add to this literature by looking at an under-explored disease, polio, in an endemic country, Pakistan.

Second, my findings contribute to the research agenda on the determinants of infectious disease incidence. Most of the existing literature has focused on studying the mistrust of vaccines (Martinez-Bravo and Stegmann 2022), the role of trade (Oster 2012), and public transportation closure (Adda 2016). In this paper, I study the role of internal displacement in spreading infectious diseases and explore some of the main transmitting mechanisms.

Finally, this paper belongs to the research agenda on the impacts of conflicts on health outcomes (Blattman and Miguel 2010; Devkota and Teijlingen 2010; Phadera 2021). I contribute to this literature by evaluating the impact of a direct consequence of conflict, forced migration, on the prevalence and transmission of polio.

3.2 Background

3.2.1 Conflict in F.A.T.A. Region

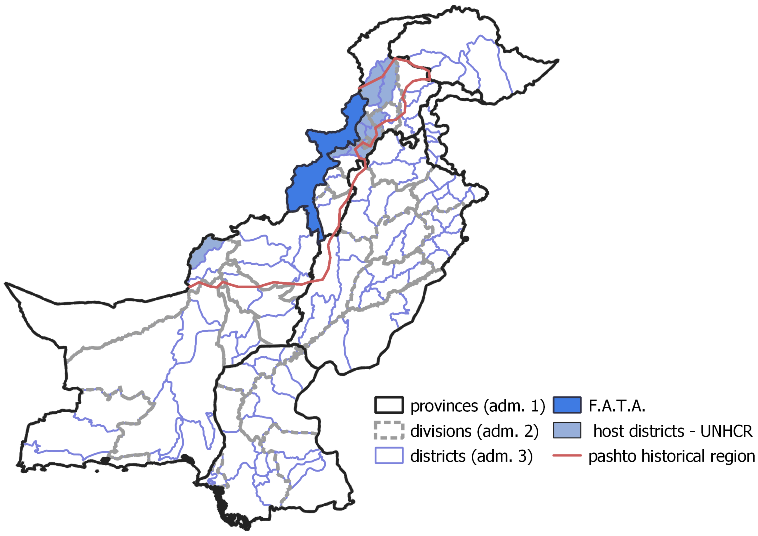
Pakistan witnessed a vast surge in violence after the terrorist attack in September 2001 in the United States (U.S.). The increase in violence manifested in waves of violent attacks against state institutions and civilians across Pakistan. Terrorists carried out around 1,600 attacks in the pre-9/11 era. However, a significant surge in the number of attacks was observed (around 12,000) in the aftermath of the 9/11 period (GTD 2021). The intensity of such violence was considerably higher in the Federally Administered Tribal Areas (F.A.T.A.) when the Tehrik-e-Taliban militants began entering into the region (Malik, Mirza, and Rehman 2023).[1]. Figure A.3.1 plots the number of attacks in Pakistan and the F.A.T.A. region.

F.A.T.A. was an autonomous tribal region in north-western Pakistan that existed from 1947 until being merged with the neighbouring province of Khyber

[1] The Pakistani Taliban, formally called the Tehreek-e-Taliban-e-Pakistan, is an umbrella organisation of various Islamist armed militant groups operating along the Afghan–Pakistani border (Abbas 2008)

Pakhtunkhwa in 2018.[2] F.A.T.A. were bordered by: Afghanistan to the north and west, Khyber Pakhtunkhwa to the east, and Balochistan to the south. Figure 3.1 illustrates the three administrative levels of Pakistan.[3] Its total population was estimated in 2000 to be about 3,341,080 people or roughly 2% of Pakistan's population, being Pakistan's most rural administrative unit. F.A.T.A. was located in *Pashtunistan* (*land of the Pashtuns* in Pashto), a historical pre-colonial region wherein Pashtun culture, the Pashto language, and Pashtun identity have been based.[4]

Figure 3.1: Host destinations and Pashtunistan



Note: This figure shows the spatial distribution of the main host districts in Pakistan. In light blue, I show the districts which received the internally displaced population from the Federally Administered Tribal Areas (F.A.T.A.) recorded by UNHCR (UNHCR 2022). The region of F.A.T.A. is in dark blue. The red line illustrates the pre-colonial region of *Pashtunistan*. Black polygons correspond to the provinces (the first administrative division in Pakistan). Grey polygons correspond to division (the second administrative division). And the white polygons with purple lines are the districts (the third administrative division).

The acceleration of violence in the F.A.T.A. led to a domestic and global policy response. After 9/11, Pakistani and U.S. forces exposed the F.A.T.A. to military offensives against alleged sanctuaries of terrorist outfits. On June 19, 2004, the U.S. undertook its first drone strike in Pakistan. Since then, the U.S. has carried out more than 406 drone attacks against alleged Al-Qaeda-linked affiliates in Pakistan's North-West. These attacks increased from 2007 and peaked around 2010. 98% of

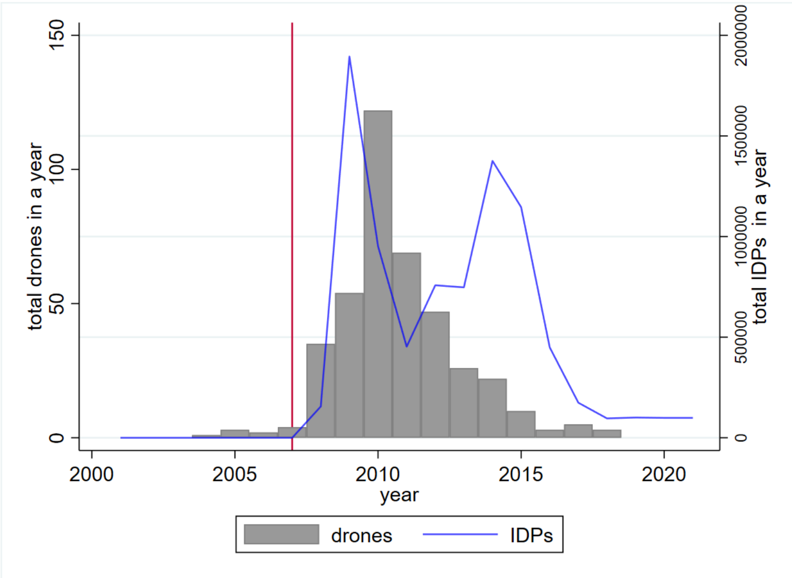
[2] The administrative units of Pakistan comprise four provinces, one federal territory, and two disputed territories: Punjab, Sindh, Khyber Pakhtunkhwa, and Balochistan; the Islamabad Capital Territory; and the administrative territories of Azad Jammu and Kashmir and Gilgit–Baltistan.

[3] A province (administrative level 1) has different divisions (administrative level 2), and a division is divided into other districts (administrative level 3).

[4] *Pashtunistan* is a historical region on the Iranian Plateau, inhabited by the indigenous Pashtun people of southern Afghanistan and north-western Pakistan. During British rule in India in 1893, Mortimer Durand drew the Durand Line, fixing the limits of the spheres of influence between the Emirate of Afghanistan and British India and dividing the historical *Pashtunistan* as a share of two different countries (Bezhan 2014).

the drone attacks were in the F.A.T.A. Figure 3.2 shows the total number of drones from 2001 to 2022 (New-America 2021).

Figure 3.2: Total drones strikes and IDP population (2000-2022)



Note: This figure shows the yearly number of drones and internally displaced persons (IDP) from 2001 to 2022. The grey bars show the number of drones and the blue line to the number of IDPs. The vertical red line corresponds to 2007.

3.2.2 Forced Displacement within Pakistan

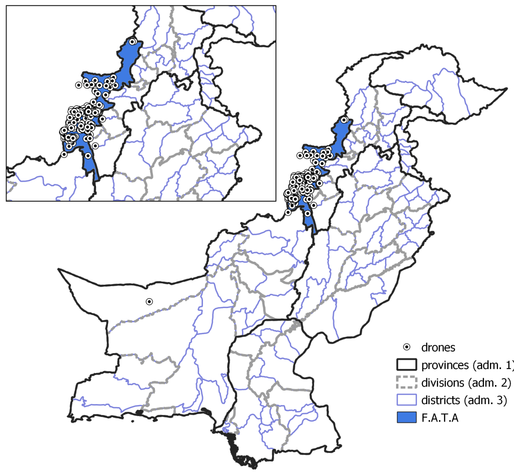
Since 2004, 13,289,880 million people have been displaced due to different operations against insurgents in F.A.T.A. Most of the affected population has been displaced multiple times after returning to their places of origin. 98% of the forcibly displaced population migrated within Pakistan. See Figure A.3.2 for a visual representation. The onset of the IDPs crises was in 2008, corresponding with a big jump in drone strikes in F.A.T.A. In 2009, the stock of internally displaced people (IDPs) reached more than 1.9 million individuals, corresponding to 57% of the F.A.T.A. population (UNHCR 2022). Figure 3.2 visually represents the total internally displaced population and the number of drones from 2001 to 2022.

The IDPs came from different F.A.T.A. districts, but especially from the most affected by the conflict (North Waziristan and South Waziristan) (UNHCR 2022). Figure 3.3 shows the positive correlation between the number of drones and IDPs from a given district. I present the total number of IDPs by origin in Table A.3.1.

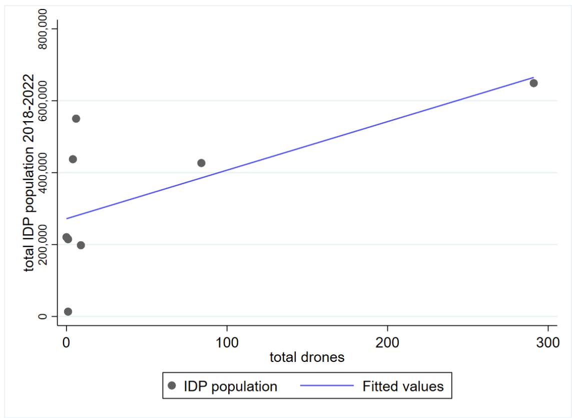
Due to cultural and linguistic similarities, most IDPs migrated to relatively safe districts within the historical region of *Pashtunistan*. Many arrived in the place of displacement as cohesive groups, which helped them maintain a sense of community. They have also utilised social networks from their home areas (IDMC 2015). IDP's

Figure 3.3: Drones as a migration push factor

a) Drone strikes locations



b) Correlation between the number of drones and IDPs



Note: This figure shows the relationship between the intensity of drone strikes and migration. Figure A illustrates the spatial distribution of drones from 2001 to 2022 in Pakistan. The blue polygons correspond to the Federally Administered Tribal Areas (F.A.T.A.) region. Figure B plots the correlation between the district's number of drones and the aggregate IDPs from 2001 to 2022.

integration in host communities was not always easy. In specific locations, IDPs were even discriminated against by the native population and political leaders whipping up xenophobia against the displaced. Such attitude forced the displaced to find shelter only among their ethnic groups, adversely affecting their ability to relocate or freely seek employment (Din 2010; IDMC 2015).

Additionally, IDPs usually reside in informal settlements in host communities and avoid living in camps for multiple reasons, including the fear of attack by non-state armed groups, poor conditions and lack of private space. Usually, the informal settlements lack safe drinking water, sanitation, and health care (IDMC 2015). I list the total IDPs in host districts by year in Table A.3.2.

3.2.3 Polio in Pakistan

Polio or Poliomyelitis is a highly infectious viral disease. The virus is transmitted by person-to-person contact. It lives in an infected person's throat and intestines. It spreads through contact with an infected person's stool (poop) or, less frequently, droplets from a sneeze. An infected person may spread the virus to others immediately before and up to two weeks after developing symptoms. The virus may live in an infected person's intestines for many weeks. They can contaminate food and water when they touch them with unwashed hands (CDC 2021).

There is no cure for polio, and it can only be prevented. Two polio vaccines are available: oral polio (OPV) and inactivated polio (IPV).[5] Children should usually get four doses of the polio vaccine at ages two months, four months, 6–18 months, and 4–6 years.

Although anyone not fully vaccinated against polio is at risk for polio, polio predominantly affects children under five. Most people who get infected with poliovirus will not have any visible symptoms. About 1 out of 4 people with poliovirus infection will have flu-like symptoms (fever, fatigue, headache, vomiting, stiffness of the neck and pain in the limbs). These symptoms usually last 2 to 5 days. One in 200 infections leads to irreversible paralysis (usually in the legs). Among those paralysed, 5–10% die when their breathing muscles become immobilised (WHO 2022).

Cases of polio have fallen dramatically over time. In 1988, more than 350,000 polio cases were reported annually across 125 countries. In 2021, the number of cases was down to 649. The main reason was the increased number of children vaccinated. Globally in 1980, only 22% of one-year-olds were vaccinated against polio, which increased to coverage of 86% of the world's one-year-olds in 2015.[6] In 2001, 14 countries reported cases of wild polioviruses. By 2021 there were only two countries where wild poliovirus cases were recorded: Afghanistan and Pakistan (WHO 2022).

Since 1994, the Pakistan Polio Eradication Programme (P.P.E.P.) has been fighting to end the crippling poliovirus in the country. In 1997, Pakistan reported 1,147 cases, constituting 22% of the cases reported worldwide. With its initial extraordinary efforts to control polio among children, Pakistan reduced cases from 20,000 in 1990 to 28 in 2005. However, about 100 cases have been reported annually after 2007. Cases steadily rose from 32 in 2007 to 118 in 2008 to 198 in 2011.

[5] OPV is administered orally and can be given by volunteers. OPV protects both the individual and the community because it induces gut immunity, which is essential to stopping poliovirus transmission. IPV is given by injection and needs to be administered by a trained health worker. IPV is highly effective in protecting individuals from severe diseases caused by poliovirus. However, it cannot stop the virus's spread in a community.

[6] In 1988, the World Health Assembly created the Global Polio Eradication Initiative to eradicate polio by 2000.

3.2.4 Inmunization in Pakistan

Children in Pakistan typically receive three primary vaccines through routine immunisation activities: the vaccine against poliomyelitis, the D.P.T. (vaccine against diphtheria, pertussis, and tetanus) vaccine, and the measles vaccine. Pakistan follows the recommended vaccination calendar of the World Health Organization, and the first dose of most vaccines is supposed to be administered shortly after birth.

As part of the P.P.E.P., Lady Health Workers are the health workers responsible for child immunisation. These workers are assigned to a local health facility, each responsible for approximately 1,000 people or 150 homes. They regularly visit households to provide family planning information and immunise children according to the vaccination schedule. Since 2010, the provision of public health goods is a provincial responsibility. In 2014, there were approximately 110,000 Lady Health Workers in Pakistan. However, the main way Pakistani children are immunised is through vaccination drives. There are national and subnational immunisation days during which vaccinators (typically lady health workers joined by other volunteers) provide vaccines at households' doorstep. They typically last for three days and target all children up to age 5 in the respective district. All the vaccines provided during immunisation drives or at public health facilities are free of charge (Martinez-Bravo and Stegmann 2022).

The surge in violence in F.A.T.A. could be one of the leading reasons behind the increase in polio cases in Pakistan. Almost 70% of Pakistan's polio cases from 2004 to 2018 were reported from this area. Unhygienic and poor sanitary conditions with large families living in packed houses resulted in widespread polio transmission. Moreover, the militants carried out continuous propaganda against polio vaccination, translating into increased vaccine refusal. As the extremists banned polio vaccination, almost 400,000 children could not be vaccinated in the tribal north during 2010–2011. Even vaccination workers began to be attacked and killed (Mushtaq et al. 2015; Rahim, Ahmad, and Abdul-Ghafar 2022).

The movement of the population during the conflict has led the P.P.E.P. to implement a particular program targeting the High-Risk Mobile Populations, or H.R.M.P.s (nomads, Internally Displaced Persons, Afghans, brick kiln workers and visiting "guest children"). The H.R.M.P. strategy requires vaccinating all eligible children at all possible opportunities, including in departing communities, transit, and communities where they settle. The P.P.E.P. vaccinates children travelling or on the move through 500 permanent transit points (P.T.P.s) across all major transit points nationwide. These P.T.P.s are set up along country and district borders and other essential transit points such as railway stations, bus stops, and highways. In 2018, P.T.P.s had vaccinated a total of 1.7 million children (UNICEF 2019).

3.3 Data

I construct a panel dataset at the district and monthly level that combines data on conflict, total forcibly displaced population, polio cases and supply-demand of vaccines.

3.3.1 Conflict data

I use two georeferenced variables to measure conflict intensity in Pakistan—the number of drone strikes and terrorist attacks at district and monthly levels.

The conflict data on drone strikes comes from the World of Drones Database developed by New America (New-America 2021). New America gathers information on each drone strike's timing (day, month, year), location (latitude and longitude) and total deaths. The World of Drones database draws upon media reports and other open-source information to track which countries and non-state actors have armed drones or are developing them; and which actors have used them in combat and where.

The New America Database has reported 406 drone strikes in Pakistan from January 1, 2001, to December 31, 2022. The first drone was recorded on June 19, 2004, and the last on July 4, 2018. Only 10 of the 406 drones were located outside F.A.T.A. Figure 3.3 presents the spatial distribution of drone strikes in Pakistan. I construct my primary measure of conflict by aggregating the drones that fall in a district monthly. To show the robustness of my results to alternative ways of measuring drone intensity, I construct a supplemental measure: the number of people killed. Figure A.3.3 presents the spatial distribution of total deaths by drone strikes in Pakistan.

The data on attacks against the state and civilians are extracted from the Global Terrorism Database - G.T.D. (GTD 2021). The G.T.D. provides details on more than 200,000 terrorist incidents worldwide since 1970. For each incident, information is provided on the timing (day, month, and year), location (latitude and longitude), fatalities (wounded and killed), type (assassination, explosion, suicide, hijacking, etc.), target (civilians, businesses, government officials, religious institutions, N.G.O.s, etc.), the terrorist group which carried out the attack, and the motivation of the episode (political or religious).

The G.T.D. reported 13,638 terrorist attacks from January 1, 2001, to December 31, 2020. I construct a measure of terrorist attacks at the district level by aggregating the number of incidents that fall in a district. I complement this measure by repeating this exercise with the number of people killed in the attacks.

3.3.2 Forced displacement data

United Nations High Commissioner for Refugees - U.N.H.C.R. provides the data on forcibly displaced populations (UNHCR 2022). UNHCR 2022 contains information about the countries of destination and origin, province and district within a country, total population, year of arrival, and demographic characteristics (age and gender). Therefore, this data allows me to identify the total internally displaced population (IDPs), total Pakistani refugees outside Pakistan, and total Afghan refugees in Pakistan. Figure A.3.2 shows how a large share of the forcibly displaced population remained within Pakistan.

Among the IDPs who fled from F.A.T.A., 54% of them are below 18 and 18% below 5. Figure A.3.4 plots the total IDP distribution by age. Moreover, 47% of the IDPs were women or girls. The destination districts were concentrated in Khyber Pakhtunkhwa, as shown in Figure A.3.5. Figure 3.1 shows the spatial distribution of the host districts reported by U.N.H.C.R., which all fall within the historical *Pashtunistan*. Figure A.3.6 plots the total IDPs in each district over time.

3.3.3 Polio data

I collect data on polio incidence from January 1, 2001, to December 31, 2022, from the Polio Eradication Program established by the World Health Organization (WHO). The Polio Eradication Program gathers information for each reported polio case on the timing (year, month and year), location (district), and the type of virus. I build my outcome measure on polio incidence by aggregating the number of new polio cases in a given district and month.

The Polio Eradication Program reported 2,080 new polio cases from 2001 to 2022. The cases in the entire country and F.AT.A. have followed a similar pattern. Figure A.3.7 shows the evolution of cases. There are three critical years where the trend switched to positive: 2008, 2012 and 2018.

3.3.4 Vaccination supply and demand

For this project, I also collect data on Pakistan's polio vaccination campaigns between 2001 and 2022. I obtain this data from the Polio Eradication Program. These data contain district-month measures of whether a polio vaccination campaign was conducted, the type of campaign—case response, mop-ups, child health days, subnational or national immunization days—, the age group targeted, and vaccine type.[7]

I rely on data from two waves of the Demographic Health Surveys (D.H.S.) in Pakistan to obtain measures of polio immunization at the individual level from 2008. In particular, I used information on the demand for the polio vaccine from the 2012/13 and 2017/18 D.H.S. surveys. Moreover, I profit from the 2006/07 and 1990/91 D.H.S. surveys to obtain measures before 2008. The D.H.S. has data on the year and month of birth, allowing me to define the exposure to the inflow of the IDP population.

The D.H.S. asks each household member whether the individual was born in the current district of residence and the reason for the migration. I exploit this migration data to build a variable on whether an individual is displaced or native in a given district. The D.H.S. also contains georeferenced household location information (only in 2006/07 and 2017/18 D.H.S. surveys).

Finally, the D.H.S. characterize the host communities at the household (i.e. sanitation, overcrowding, house conditions, and health provision) and individual level (i.e. health-seeking behaviour, labour, and education).

3.3.5 Other data

For my empirical strategy, I identify the historical pre-colonial region of *Pashtunistan* in Pakistan from the Georeferencing Ethnic Power Relations - GeoEPR 2021 dataset (Vogt et al. 2015). GeoEPR geo-codes all politically relevant ethnic groups from the Ethnic Power Relations-Core 2021 dataset and provides polygons describing their location on a digital map.

To test for the validity of my identification strategy, I use additional controls, including constructed district-year level data on satellite night light density as a proxy of economic development. The National Oceanic and Atmospheric Administration (NOAA) processes night light density data. NOAA uses satellite images collected by the U.S. Air Force Defense Meteorological Satellite Program. Two satellites that circle the Earth 14 times daily collect the images, recording the intensity of Earth-

[7] Mop-ups are very targeted, geographically limited polio campaigns, held between larger-scale national Immunization Days or subnational Immunization Days in areas where we know many children were missing, for example, or where a large immunity gap persists. Child Health Days are not specifically polio campaigns, but the polio vaccine is added to Child Health Days alongside other vaccines and health interventions. National Immunization Days are nationwide campaigns targeting all children aged 0-5. Subnational Immunization Days are vaccination campaigns in key high-risk provinces.

based lights with their Operational Linescan System. I also use sociodemographic data from the 1998, 1981, and 1973 Population Census.[8]

3.4 Identification Strategy

This paper aims to study the impacts of inflows of internally displaced populations on the hosting communities' polio incidence. To tackle this question, my identification strategy relies on comparing district new polio cases in locations exposed to large IDPs inflows with new polio cases in those minor or non-affected before and after the onset of the IDP crises of 2008.

IDPs settlement is a potential endogenous decision, and time-varying characteristics in host communities could affect the resettlement pattern and polio incidence. For example, IDPs might choose to migrate to poor areas closer to their original communities, leading us to overestimate the harmful effects of IDP population on the number of polio cases. Additionally, most IDPs settle down in host communities without registration systems (UNHCR 2017). Herefore, the officially identified host districts and the total IDPs in each community may be underestimated. To overcome these challenges, I exploit the proximity to the *Pashtunistan* historical border to define the treatment and counterfactual.

3.4.1 Pashtunistan historical border

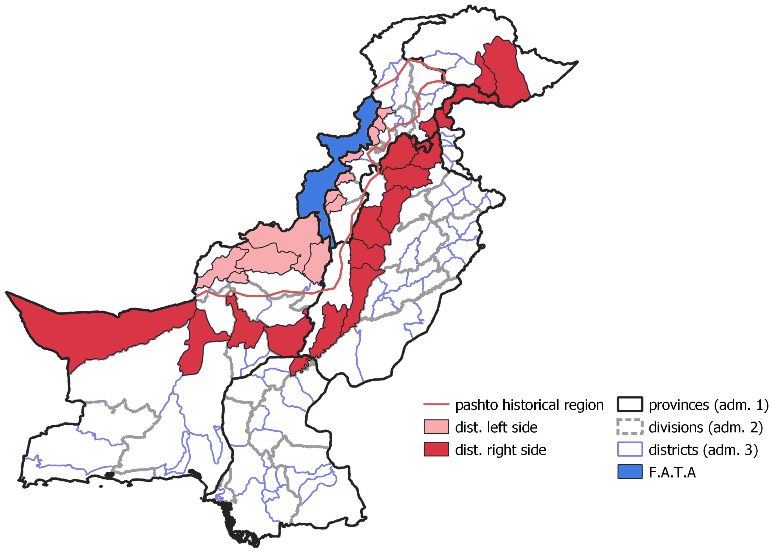
F.A.T.A. is part of the pre-colonial region of *Pashtunistan*. And, due to cultural and linguistic similarities, many of the IDP families from F.A.T.A. migrated to other districts within *Pashtunistan*, as shown in Figure 3.1. The border between Afghanistan and Pakistan (known as *the Durand Line*) results from an agreement in 1893 between the British Indian government and the emir of Afghanistan. Eighty-five per cent of the Durand Line follows rivers and other physical features, not ethnic boundaries, splitting the historical Pashtu region into two separate countries.

The native people of *Pashtunistan* are the Pashtuns. They are the largest ethnic group in Afghanistan and the second largest in Pakistan. The main language spoken in the delineated *Pashtunistan* region is Pashto. The Pashtuns practice Pashtunwali, the indigenous culture of the Pashtuns. This pre-Islamic identity remains significant for many Pashtuns and is one factor that has kept the *Pashtunistan* culture alive. Although the Pashtuns are politically separated by the Durand Line and other administrative borders within Pakistan, Pashtun tribes tend to ignore the borders.

[8] The 1998 Census sample covers a share of Khyber Pakhtunkhwa (23 out of the 31 districts) and Punjab (24 out of the 38 districts). The 1973 Census sample covers all Balochistan (28 districts) and Punjab (38 districts), a shared of Khyber Pakhtunkhwa (18 out of the 31 districts) and Sind (26 out of the 27 districts). The 1981 Census has information for 76 out of the 141 districts in Pakistan.

For instance, many Pashtun tribes from the F.A.T.A area and the adjacent regions of Afghanistan used to cross back and forth with relative ease to attend weddings and other events. After 2004, this cross-border movement is checked via the military and has become much less common compared to the past. However, the transit across *Pashtunistan* districts in Pakistan has never stopped, allowing IDPs to move within their historical region.

Figure 3.4: Treated and control districts



Note: This figure shows treated (host) and control (non-host) districts. To define them, I use the spatial distribution of districts relative to the pre-colonial region of *Pashtunistan*. The red line corresponds to the *Pashtunistan*'s border. Districts whose territory falls within the pre-colonial region of *Pashtunistan* are host districts. Non-host districts are those whose territory is outside *Pashtunistan* but adjacent to the historical border.

To identify the effects, I compare the polio cases in districts within *Pashtunistan* (treatment) with those in districts immediately outside *Pashtunistan* (counterfactual) before and after 2007. Figure 3.4 shows the Pakistani districts within and outside Pashtunistan. The central identifying assumption is that the IDP population mostly moved to Pashtu districts, and non-Pashtu districts had no or negligible presence of IDPs. Nevertheless, there are two features worthy of highlighting. First, there are some districts in which only a share of their territory is within *Pashtunistan*. I start by removing them from the sample since I can not define them as being within or outside *Pashtunistan*. Second, *Pashtunistan* covers F.A.T.A., but I drop it from the sample to avoid potential confounding conflict effects. Hence, the treated districts are those on the left side of the *Pashtunistan* border and the control districts on the right. The design of my counterfactual allows me to compare host districts (Pashtu districts) to the most similar administrative units possible (the closest non-Pashtu districts). I use alternative definitions of treated and control districts in the robustness section. See Figure 3.4 for a spatial distribution of the treatment and counterfactual.

I estimate the following specification:

$$Y_{d,p,t,m} = \beta_1 Pashtunistan_d \ \ X \ \ IDP \ \ Crises_t + \beta_2 X_d + \alpha_p + \gamma_{t,m} + \epsilon_d \tag{3.1}$$

where d stands for district, p stands for province, t for year and m for month. $Y_{d,p,t,m}$ represents the district outcome: one if at least a polio case in the year and month t-m, zero otherwise. $Pashtunistan_d$ stands for being or not within *Pashtunistan*: one if district d falls in the historical region, zero otherwise. $IDP \ \ Crises_t$ is a dummy variable that takes the value of one after 2007. X_d is a matrix of district-year controls. Namely, I control for district-level nightlight intensity as a proxy of economic development (Pérez-Sindín, Chen, and Prishchepov 2021) and the total number of polio vaccination campaigns in a given district month which account for the vaccination supply. α_p and $\gamma_{t,m}$ account for province and year-month fixed effects. Standard errors are clustered at the district level to account for time serial correlation in the outcome across geographic areas. A battery of robustness tests that support the validity of my identification strategy is presented in sections 5.2 and 7.

3.5 Results

I first examine whether there are different likelihoods of having at least one polio case in IDP population host districts to districts with no or negligible presence of IDP. The specification estimates in equation (1) are presented in Panel A of Table 3.1.

The results suggest that districts that received the IDP population have experienced an increase in the probability of polio incidence compared to non-host districts. Column 1 shows the estimates without fixed effects and controls. An IDP inflow increased the probability of at least one polio case in the host districts by 4.1 percentage points (pp). The estimates are statistically significant at the one per cent level. The magnitude of the effects does not change when adding province-fixed effects and year-month fixed effects (column (2)), but the estimates are significant at the five per cent level. Column (3) shows that the point estimates increase to 5.3 pp when controlling for nightlight intensity and total vaccination campaigns (significant at the five per cent level). We could be concerned that the different characteristics in sanitation and overcrowding between treated and control districts could drive my results.

In column (4), I control for the average number of children under five, the average number of members in a household, and the total share of the literate population in a district from the 1973, 1981 and 1998 Population Census. The results hold and are

Table 3.1: Effect of IDP population on polio

	(1)	(2)	(3)	(4)	(5)
	PANEL A: At least one polio case				
VARIABLES	polio	polio	polio	polio	polio
2007 x Host district	0.038***	0.038**	0.041**	0.050**	0.054**
	(0.009)	(0.014)	(0.017)	(0.021)	(0.023)
Observations	8,184	8,184	8,184	6,600	6,060
Number of districts		31	31	25	31
	PANEL B: Number of polio cases				
VARIABLES	polio cases	polio cases	polio cases	polio cases	polio cases
2007 x Host district	0.081***	0.081**	0.081**	0.093**	0.111**
	(0.015)	(0.032)	(0.034)	(0.039)	(0.048)
Observations	8,184	8,184	8,184	6,600	6,060
Number of districts		31	31	25	31
	PANEL C: Number of polio cases per 100,000 inhabitants				
VARIABLES	polio pop.den.	polio pop.den.	polio pop.den.	polio pop.den.	polio pop.den.
2007 x Host district	0.006575***	0.006575**	0.006661*	0.007462*	0.009270**
	(0.001872)	(0.003130)	(0.003444)	(0.003901)	(0.004001)
Observations	7,128	7,128	7,128	6,600	5,760
Number of districts		27	27	25	27
Province FE	No	Yes	Yes	Yes	Yes
Time FE	No	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes

Note: This Table presents the impacts of the IDP inflows on district polio prevalence in host districts compared to non-host district. I use the spatial distribution of districts with respect the pre-colonial region of *Pashtunistan* to define host and non-host districts. Districts whose territory fall within the pre-colonial region of *Pashtunistan* are defined as host districts. Non-host districts are the district whose territory is outside *Pashtunistan*, but are adjacent to the historical border. The treatment timing starts from 2008. Observations are at the district and month level from 2001 to 2022. The baseline specification is presented in equation (1). Column (1) presents the results without province, year-month fixed effects and covariates. Column (2) includes province and year-month fixed effects. Column (3) controls for nightlight intensity and total vaccination campaigns. Columns (4) controls for pre-treatment district-covariates (the average number of children under five, the average number of members in a household, and the total share of the literate population from 1973, 1981 and 1998 Population Census). Column (5) control instead for contemporary characteristics (the average number of children under five, the average number of members in a household, shared households with piped water, and shared households with a finished floor). This Table present three different outcomes: at least one case of polio (panel A), total number of polio cases (panel B) and polio cases per 100,000 inhabitants (panel C). Standard errors are clustered at the district level. Robust standard errors are in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

significant at the 10 per cent level. The issue with the population census is that none of the censuses covers the entire sample. So, to increase the coverage, and as a result the sample, I combine them. However, it implies that the covariates are measured at different points in time. In column (5), I control for contemporary characteristics instead (the average number of children under five, the average number of members in a household, and shared households with piped water). The magnitude of the effects slightly increases and is significant at the five per cent level. When I look at the absolute number of polio cases in Panel B of Table 3.1, the results hold.

3.5.1 Contextualizing the Magnitude of the Effects

What is the impact on polio cases relative to the population size? I obtain similar results when estimating equation (1) with a continuous outcome: the number of new polio cases per 100,000 inhabitants. I present the results in panel C of Table 3.1. I use the population in 2017 to calculate the number of cases per 100,000 inhabitants.[9] An inflow of the IDP population results in 0.007 additional cases per 100,000 inhabitants (column (3)). Although negligible, this represents an increase of 40% over the mean incidence.[10] The intuition behind this finding is that host districts are more likely to encounter a polio case and to experience a more rapid spread of the virus.

Additionally, I use the population in 1998 to estimate the polio cases per 100,000 native inhabitants. Table A.3.3 shows that an inflow of the IDP population corresponds with 0.011 additional cases per 100,000 native inhabitants (see column (3)). The points estimates are statistically significant at the five per cent level.

Does the intensity of the IDP inflow affect the results? Precise data on the IDP inflow at the district-year level does not exist for my entire timeframe. Therefore, I approximate district-year inflows of the IDP population using the following measure:

$$PredictedInflow_{dpt} = IDPInflow_t \text{ x } \frac{1}{distance_{dp}}$$

where $IDPInflow_t$ represents the total inflows of the internally displaced population registered in Pakistan in each year t, and $distance_{dp}$ is the Euclidean distance from the centroid of each district d from the province p to F.A.T.A. border. I construct my predicted inflows measure as the interaction of the inverse distance of each district to the closest F.A.T.A. border (district variation) and the total yearly number of IDP population (annual variation).

The distribution of $PredictedInflow_{dpt}$ across the inverse distance of each district to the F.A.T.A. border is displayed in Figure A.3.8. To test whether this cross-section variation is associated with IDP migration patterns, I use the available data on the IDP population at the district level from the U.N.H.C.R. Figure A.3.9 shows that my inverse distance measure correlates highly with the reported number of IDP populations by U.N.H.C.R.

Panel A in Table 3.2 shows the equation (1) results with continuous treatment, $PredictedInflow_{dpt}$. The intensity of the IDP inflow has a significant effect on polio incidence. An increase of one standard deviation in predicted inflows results in 0.001 additional cases per 100,000 inhabitants (see Panel C in Table 3.2), corresponding to a 20% of the mean incidence.

Why are the main effects meaningful? Eradicating polio has been a worldwide effort over the last decades. But, until polio is completely eradicated, all countries remain at risk of imported wild poliovirus. Identifying the determinants of new cases is critical to prevent additional ones. Polio, which only infects humans, has been eliminated in 193 countries. With the transmission of wild-type polio limited

[9] Azad Jammu and Kashmir and Gilgit-Baltistan provinces are not in the 2017 Population census. So, Kargil, Kupwara, Muzaffarabad and Neelum districts drop from the sample.

[10] The average number of new polio cases in my sample from 2001 to 2022 equals 0.005.

Table 3.2: Effect of predicted IDP inflow on polio

	(1)	(2)	(3)	(4)	(5)
			PANEL A: At least one polio case		
VARIABLES	polio	polio	polio	polio	polio
Predicted Inflow	0.022901***	0.015548*	0.015931*	0.018607*	0.017538*
	(0.003250)	(0.008464)	(0.008851)	(0.010282)	(0.009651)
Observations	8,184	8,184	8,184	6,600	6,060
Number of districts		31	31	25	31
			PANEL B: Number of polio cases		
VARIABLES	polio cases	polio cases	polio cases	polio cases	polio cases
Predicted Inflow	0.041253***	0.032000	0.031959	0.036553	0.036966
	(0.007991)	(0.019921)	(0.020238)	(0.023178)	(0.022757)
Observations	8,184	8,184	8,184	6,600	6,060
Number of districts		31	31	25	31
			PANEL C: Number of polio cases per 100,000 inhabitants		
VARIABLES	polio pop.den.	polio pop.den.	polio pop.den.	polio pop.den.	polio pop.den.
Predicted Inflow	0.001876***	0.001423***	0.001403**	0.001621**	0.001287*
	(0.000470)	(0.000503)	(0.000556)	(0.000634)	(0.000737)
Observations	7,128	7,128	7,128	6,600	4,740
Number of districts		27	27	25	27
Province FE	No	Yes	Yes	Yes	Yes
Time FE	No	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes

Note: This Table presents the results on polio prevalence with a continuous treatment: the predicted IDP inflows. I construct my predicted inflows measure as the interaction of the inverse distance of each district to the closest F.A.T.A border (district variation) and the total yearly number of IDP population (annual variation). The treatment timing starts from 2008. Observations are at the district and month level from 2001 to 2022. The baseline specification is presented in equation (1). Column (1) presents the results without province, year-month fixed effects and covariates. I restrict the sample to the districts defined as host or non-host district. Column (2) includes province and year-month fixed effects. Column (3) controls for nightlight intensity and total vaccination campaigns. Columns (4) controls for pre-treatment district-covariates (the average number of children under five, the average number of members in a household, and the total share of the literate population from 1973, 1981 and 1998 Population Census). Column (5) control instead for contemporary characteristics (the average number of children under five, the average number of members in a household, shared households with piped water, and shared households with a finished floor). This Table present three different outcomes: at least one case of polio (panel A), total number of polio cases (panel B) and polio cases per 100,000 inhabitants (panel C). Standard errors are clustered at the district level. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

to Afghanistan and Pakistan, an official eradication declaration is in sight. In 2005, 28 cases were reported in Pakistan, compared to the 1,147 cases in 1997. Moreover, most host districts had zero or close to zero polio cases before 2007. Ultimately, the estimates I present in this paper capture the impacts of IDP inflows on keeping host communities away from eradicating polio rather than the effects of an extensive increase in polio cases. Even if these results are substantial in magnitude. A high risk of underreporting is present, as 75% of people infected with poliovirus are asymptomatic (WHO 2022). Ideally, I could look at the incidence of other diseases, such as measles, chickenpox, or malaria, to validate my findings. Unfortunately, I could not find comparable data for other diseases.[11]

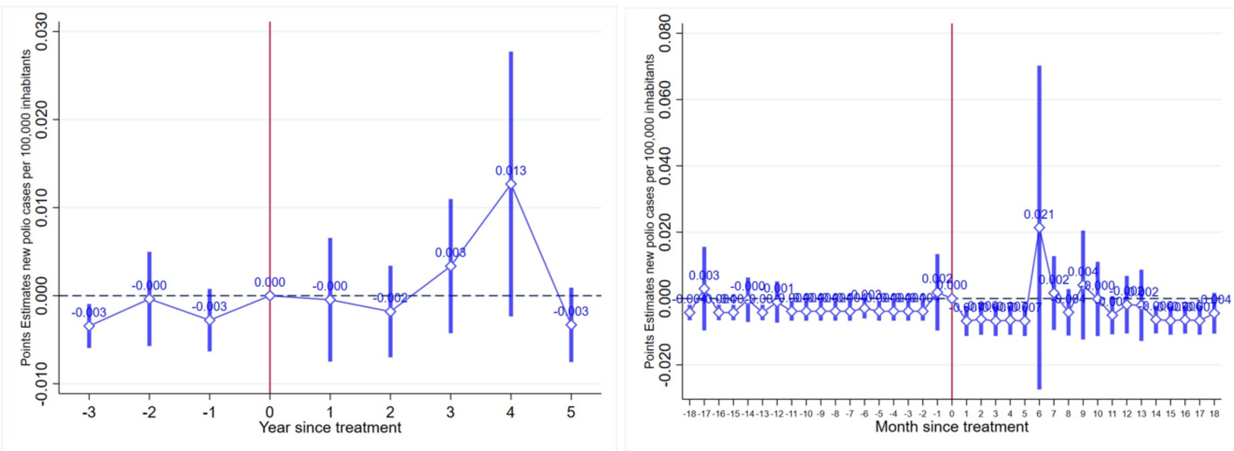
[11] The information on new polio cases from 2001 to 2022 is not publicly available. I obtained this information from the Expanded Program of Immunization. The data available for the other diseases were insufficient to grant a meaningful regression analysis or were not available before 2008.

3.5.2 Threats

Unbalanced pre-treatment characteristics. Table A.3.4 presents summary statistics of baseline key demographics and socioeconomic variables that compare districts with high IDP intensity ($Pashtunistan_d = 1$) and districts without or with negligible IDP population ($Pashtunistan_d = 0$). The balancing tests do not reveal significant pre-shock differences except that treated districts seem to have a higher share of households with piped water, toilets, television, and in urban locations, and an additional number of children under five and four more members in the household on average terms. However, by definition, the regions may have differences in social beliefs and norms. Since in host districts, Pashtu traditions and norms are predominant in host districts. Table A.16, in the robustness section, shows similar results with a counterfactual within the pre-colonial region of *Pashtunistan*, ruling out the possibility that social norms differences could bias the main findings.

Pre-trends outcomes. The findings indicate that there are no pre-trends in the outcomes. The key identifying assumption of the main results is that host and non-host districts should evolve similarly without treatment. In other words, there may not exist differences in the pre-treatment outcome between treated and control districts. Figure 3.5 estimates an event study with the new cases per 100,000 inhabitants, suggesting that non-consistent differences exist. I generate additional evidence in Figure A.3.10 by plotting the new polio cases from 2000 to 2022 in treated and host districts.

Figure 3.5: Parallel trend test: Event study



Note: Figure 3.5 plots the event and year coefficient from estimating equation 1 using the new polio cases per inhabitant as the dependent variable. The confidence intervals are 95%. Polio outcomes come from the Polio Eradication Program established by the World Health Organization (WHO). The omitted category is T-0, the year 2007. The dataset is in a year-district panel format. Treatment is defined at the year level. On the right-side diagram of this Figure, I repeat the exercise by month from the treatment.

In baseline specification 3.1, I include province-fixed effects. We could be concerned about the heterogeneity of groups. In Figure 3.25, I conduct the same exercise as in Figure 3.5 but by treated provinces. Even if looking at the results by treated province, the parallel trend assumption holds when estimating an event study at the monthly level.

Conflict effect. Conflict can affect the health outcomes of children at early age (Bundervoet, Verwimp, and Akresh 2009). The conflict is primarily concentrated in F.A.T.A., which is not in my baseline sample. Still, we could be concerned about potential spillovers in the neighbouring district to F.A.T.A., which could bias my findings. The 86% of the drone strikes were located on the southern divisions of F.A.T.A. (in North Waziristan, South Waziristan and Bhittani districts). We could assume that adjacent districts to F.A.T.A.'s northern districts were less indirectly exposed to the conflict. Table A.3.5 shows in panels A and B that the results of Table 3.1 hold at a certain extend when I restrict my treatment and counterfactual to the neighbouring northern or southern districts.[12]

Terrorist attacks took place across Pakistani districts. In particular, 43% of the districts of my sample experienced at least a terrorist attack during my period of analysis (2001 to 2022). The intensity of the attacks could contaminate my estimates. I control the number of attacks in a given district to rule out this hypothesis. The points estimates remain statistically significant, but the point estimates slightly decrease to 0.044 (see panel C of Table A.3.5).

Afghan refugees. Since the late 1970s, Pakistan has been a host country for millions of refugees and some 1.35 million still reside in the country (UNHCR 2022). Figure A.3.11 shows the evolution of total Afghan refugees in Pakistan from 2001 to 2022. Most refugees are in the Pashtun-dominated areas of Pakistan. This fact is a major problem for my identification. To upfront this empirical limitation, I conduct three different exercises. First, I show that the results of Table 3.1 hold when I control for the total district-year refugees (see panel A of Table A.3.6). Only 3 out of the 296 camps had IDPs as the targeted population. It made that most of the refugees live in refugee camps. Thus, as a second exercise, I also show evidence that the estimates do not change when I control for the number of camps in a district. Finally, the results do not substantially change when I add an interaction to the number of camps in a district. Importantly, the interaction is statistically non-significant, providing suggestive evidence that the effects are not driven by the presence of Afghan refugees. See the results in panels B and C of Table A.3.6. The 67% of the treated districts have a refugee camp compared to the 26% of the non-

[12] The districts in the North are: Abbottabad, Attok, Chakwal, Charsadda, Hangu, Islamabad, Kargil, Kupwara (Gilgit Wazarat), Malakand P.A., Muzaffarabad, Neelum, Peshawar, and Rawalpindi. But only 4 of these districts are treated, which could explain the loose of power in Columns (3), (4), and (5). Panel B shows the estimates for a sample of Southern districts (Bannu, Bhakkar, Bhittani, Bolan, Chagai, Dera Bugti, Kalat, Kashmore, Khushab, Layyah, Musakhel, Muzaffargarh, Pishin, Qilla Saifullah, Rajan Pur, Tank, Zhob, and Ziarat).

host districts. So, even if all these checks hold, I can not rule out completely that my estimates do not capture the effects of refugees itself.

Migration outflows. Although very few Pakistanis migrated internationally, a big jump in the number of Pakistani refugees before and after 2007 could affect my results. Figure A.3.2 helps to remove this concern. The number of Pakistani refugees has been relatively constant from 2000 to 2011, with an increase from 2012. However, the results remain unchanged when I control for the total number of refugees from Pakistan in each year (see Table A.3.7).

3.6 Mechanisms

There are three channels through which IDP inflows could increase polio incidence in host communities. First, low vaccination rates at the national level increase the susceptibility to transmitting polio in overpopulated communities. Second, the precarious health conditions in host communities may facilitate the spread of the virus. Third, a sudden increase in the population could congest the health services in host communities. Yet, one caveat of the identification strategy implemented in this paper is that it cannot disentangle the precise mediating mechanisms underlying the observed increase in the number of new polio cases reported in treated communities with respect to control communities. Notwithstanding this shortcoming, in this section, I provide empirical evidence below that, although they are not conclusive, the evidence suggests that the three proposed mechanisms could have a role in the increase in polio incidence. I use individual-level data from the Demographic and Health Survey from 1990 to 2017 to analyze the mechanisms.

3.6.1 Overpopulated communities with low inmunization rates

Pakistan is one of the most populous countries in the world and one of the least developed, with a large population of approximately 188.9 million people, including 24.7 million children under five years old (UNDESA 2015). Pakistan's sociodemographic characteristics make polio eradication critical. Conflict and insecurity affecting routine immunisation teams' access generated a decline in immunisation coverage to less than 45% in F.A.T.A. (Hussain et al. 2016). As a result, we could expect that the massive arrival of the population from the F.A.T.A. decreased the immunisation rates in the host communities, increasing the number of new polio cases. To test this hypothesis, I use individual-level information on vaccination against polio and the date of birth from the Demographic and Health Survey (D.H.S.). First, Figure A.3.12 shows no differences in the share of children vaccinated against polio with at least one dose between treated and control districts after 2007. Second, I estimate the following specification to exploit within-district cohort variation:

$$Y_{i,d,k} = \beta_1 Pashtunistan_d \ \ X \ \ IDP \ \ Crises_k + \beta_2 X_i + \alpha_d + \gamma_k + \epsilon_d \tag{3.2}$$

where $Y_{i,d,k}$ is equal to one if child i from the cohort k living in district d received at least one dose of the polio vaccine, zero otherwise. The timing of the treatment is given by the year and month of birth: $Crises_k$. $Crises_k$ is one if child i was born after December 2007. I control for timing covariates at the district and individual levels. District-level covariates include nightlight intensity and the number of polio activities in the year of the interview. And I include the work status of the head of household, urban location, and gender of the child as individual-level covariates. Additionally, I include district α_d and cohort γ_k fixed effect, which accounts for generational changes in immunisation supply or social patterns.

I find that children in host communities born after the arrival of the IDP population are less likely to be vaccinated than those born before. Panel A of Table 3.3 shows the results. It seems that the IDP inflow decreases the probability of immunisation in the host districts by 6.4 percentage points (pp) in Column (2). The estimates are statistically significant at the ten per cent level. The magnitude of the effects does not change with covariates. Column (3) shows that the point estimates increase to 8 pp when controlling for nightlight intensity and the total vaccination campaigns in a year (significant at the ten per cent level). We could be concerned that the employment or sociodemographic characteristics could drive my results. In column (4), I control for the head of household's work status, urban location, and gender of the child. The results hold and are significant at the five per cent level. The results also hold when I simultaneously control for the covariates of columns (3) and (4) (see column (5)). These findings suggest that decreasing immunisation rates over time could have affected the number of polio cases.

One caveat of the above-results is that I can not disentangle if the inflow of the IDP population drives the decrease in polio immunisation in host districts. Similar changes may happen in the non-host districts. Therefore, to account for the spatial variation between host and non-host communities, I now change the district-fixed effects of equation (2) for province-fixed effects. For this new specification, there are no differences between treated and host districts in the share of children vaccinated after 2007 (see Table A.3.8). Figure A.3.12 supports this finding. We can observe how the share of children vaccinated is pretty similar between host and non-host districts over time, with a decrease around 2007.

Then, how could immunization rates affect the results? In some instances, the population of villages and towns doubled within a brief timeframe with the influx of the IDP population. The vaccination rates after 2007 barely reached 40% in both

Table 3.3: Effect of IDP population on vaccination against polio

	(1)	(2)	(3)	(4)	(5)
VARIABLES	vaccinated	vaccinated	vaccinated	vaccinated	vaccinated
			PANEL A: Cohort variation		
2007 x Host district	0.004	-0.064*	-0.080*	-0.067**	-0.083*
	(0.011)	(0.033)	(0.041)	(0.032)	(0.041)
			PANEL B: Heterogenity between IDP and native children		
2007 x Host district	0.004	-0.064*	-0.080*	-0.067**	-0.082*
	(0.011)	(0.033)	(0.041)	(0.031)	(0.041)
2007 x Host district x IDP	-0.004	-0.041	-0.041	-0.046	-0.046
	(0.075)	(0.109)	(0.109)	(0.109)	(0.109)
Observations	10,608	10,608	10,608	10,563	10,563
Province FE	No	Yes	Yes	Yes	Yes
Cohort FE	No	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes
Number of districts		31	31	31	31
			PANEL C: Districts variation		
Predicted Inflow	0.021***	-0.004	-0.005	-0.012	-0.011
	(0.005)	(0.015)	(0.016)	(0.017)	(0.017)
Observations	10,608	10,608	10,608	10,563	10,563
Province FE	No	Yes	Yes	Yes	Yes
Cohort FE	No	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes
Number of provinces		7	7	7	7

Note: This Table presents the impacts of the IDP inflows on vaccination behaviours at individual level. The dependent variable is a binary variable for being vaccinated, coded to one if the children is vaccinated. I use individual-level data on vaccination against polio and the date of birth from the Demographic and Health Survey (DHS) from 1998 to 2017. Children born from January 2008 are exposed to the treatment. The baseline specification is presented in equation (2), where I rely on within-district cohort variation. Column (1) presents the results without district, cohort fixed effects and covariates. Column (2) includes district and cohort fixed effects. Column (3) controls for nightlight intensity and total vaccination campaigns. Columns (4) controls for the head of household, urban location, and gender of the child. Column (5) control for the full set of covariates included in columns (3) and (4). Standard errors are clustered at the district level. Panel A present the baseline results. Panel B shows the estimates adding an interaction if a child is IDP. Panel C control for province fixed effects to compare within-cohort children between host and non-host districts. I use the spatial distribution of districts with respect the pre-colonial region of *Pashtunistan* to define host and non-host districts. Districts whose territory fall within the pre-colonial region of *Pashtunistan* are defined as host districts. Non-host districts are the district whose territory is outside *Pashtunistan*, but are adjacent to the historical border. Robust standard errors are in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

treated and control districts. Hence, such low immunisation rates in overpopulated host communities facilitate polio transmission. This hypothesis is very in line with the increase in polio incidance or Table 3.1.

Are the IDP children less likely to be vaccinated? I generate three pieces of evidence suggesting no statistical differences between IDP and native children. First, I add an interaction dummy in equation (2), being one if child i is internally displaced, 0 otherwise. Panel B of Table 3.3 shows that the interaction is not statistically significant. In the 2012 and 2017 waves of the D.H.S., there is a self-reported question on the reasons to migrate with an answer on "escape war or violence". I use this information to build my dummy variable. However, we need to assume a certain level of underreporting, which could affect the reliability of the results. Second, we would expect that if IDP children are less likely to be vaccinated, a higher intensity in the number of IDPs higher the aggregate number of non-vaccinated children. In Panel C of Table 3.3, I repeat the exercise of Panel A with the predicted inflow. The points estimates are statistically non-significant. Third, I conduct the same analysis of panel A of Table A.3.8, including province-fixed effects and the predicted inflow in panel B. The results are non-significant, either.

Although the vaccination of children decreased strongly in F.A.T.A. from the beginning of the conflict, a vaccination programme targetting children on the move could be the main reason behind the above results. For Pakistan's polio programme, the High-Risk Mobile Populations, or H.R.M.P.s– nomads, Internally Displaced People, Afghans, brick kiln workers and visiting "guest children"– are critical. The H.R.M.P. strategy requires vaccinating all eligible children at all possible opportunities, including in departing communities, transit, and communities where they settle. The Pakistan Polio Eradication Programme vaccinates children travelling or on the move through 500 permanent transit points (P.T.P.s) across all major transit points nationwide. These P.T.P.s are set up along country and district borders and other essential transit points such as railway stations, bus stops, and highways. The programme has developed impressive strength for vaccinating H.R.M.P. on the move. In 2018, P.T.P.s had vaccinated a total of 1.7 million children. For instance, the National Emergency Operations Centres of Pakistan have vaccinated children under ten at major transit points-border areas in southern Khyber Pakhtunkhwa, as this area has been home to some of the more high-risk mobile populations. Figure A.3.13 shows the location of the P.T.P.s surrounding the south of the F.A.T.A. region (UNICEF 2019).

In sum, one major channel of how an IDP inflow could affect the results may be the sudden increase in the population size in host communities. In a country with low immunisation rates, overpopulated communities could become suitable locations for new polio cases.

3.6.2 Poor Conditions in Host Communities

Many IDP families migrated to informal settlements, Pashtun slums or were squeezed into the houses of friends or relatives. Access to safe drinking water and hygiene is a significant problem for them. Appropriate facilities for bathing, doing laundry or keeping personal hygiene are not always available, facilitating the transmission of polio (IDMC 2015).

One crucial question is whether IDP settle in poorer locations or if the living conditions get worst with the sudden arrival of new population. The results of Table A.3.9 suggest that IDP population move to the poorest locations. Table A.3.9 shows how the number of household members and children under five was larger in host districts than in non-host districts before 2008. Moreover, Table A.3.9 presents evidence that households in host districts were also more likely to live in urban settings and less likely to have a head of household working before 2008. These pre-treatment characteristics may be a key channel behind the main results and, as well, a vital identification threat. Even if I control for local economic development, I can not ensure that my estimates capture the actual impact of IDP inflows rather than the pre-treatment differences in disadvantages characteristics. Nonetheless, what is certain is that poorer communities cannot respond efficiently to an IDP inflow, which implies that they are systematically more affected by the waves of displaced persons. Keeping this limitation in mind, how does the arrival of IDPs affect the local health conditions? To shed light on this question, I estimate equation (3) on six outcomes related to household conditions in host communities (i.e. access to drinkable water, access to a toilet, floor quality, number of children under five, households member, and head of the household's working status).

$$Y_{h,d,t} = \beta_1 Pashtunistan_d \;\; X \;\; IDP \;\; Crises_t + \beta_2 X_h + \alpha_d + \gamma_t + \epsilon_d \tag{3.3}$$

where $Y_{i,d,t}$ is equal to one if household h living in district d has a given household characteristic at the time of the survey t. The timing of the treatment is given by the year of the survey: $Crises_t$. $Crises_t$ is one if household h was interviewd after December 2007. I control for nightlight intensity and urban location. Additionally, I include district α_d and time of survey γ_t fixed effect, which accounts for time-invariant covariates across households in a district.

I find that households exposed to the IDP inflow decrease the probability of having a piped water system compared to non-exposed households (see Panel A of Table 3.4 for the results). However, I do not observe an increase in cases of diarrhoea or fever. However, that is not the case (see Table A.3.10). I obtain the same results with the predicted inflow measure (see panel B of Table 3.4). In panel C, I look at the heterogeneity between IDP and native families. IDP households are less likely to have additional children and a head of household working but more likely to have an additional household member and piped water.

The above results suggest that IDPs settle in communities with poorer conditions where a new polio case can quickly multiply. Still, the quality of the communities deteriorates with the arrival of the IDP population, with the IDPs as the most affected.

3.6.3 Congested Health Services

An alternative hypothesis could be that the displaced families created logistical hurdles in delivering subsistence and healthcare assistance to the scattered communities. Furthermore, the increased demand for healthcare services could have caused additional strain on the local infrastructure, which was often hardly adequate even

Table 3.4: Effect of IDP population on household health conditions

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	water piped	toilet	floor	total children	total members	working
	PANEL A: Cohort variation					
2007 x Host district	-0.221***	-0.148	-0.089	-0.271	-0.461	-0.004
	(0.072)	(0.091)	(0.075)	(0.248)	(0.886)	(0.039)
	PANEL B: Cohort variation - predicted inflow					
Predicted Inflow	-0.093***	0.003	0.005	-0.051	0.015	-0.009
	(0.017)	(0.023)	(0.031)	(0.056)	(0.205)	(0.015)
	PANEL C: Heterogenity between IDP and native children					
Predicted Inflow	-0.095***	0.004	0.005	-0.047	0.002	-0.008
	(0.017)	(0.023)	(0.031)	(0.057)	(0.208)	(0.015)
Predicted Inflow x IDP	0.058**	-0.031	-0.022	-0.130*	0.434*	-0.023***
	(0.022)	(0.030)	(0.032)	(0.072)	(0.229)	(0.006)
Observations	10,623	10,623	7,865	10,623	10,623	10,578
Number of districts	31	31	31	31	31	31

Note: This Table presents the impacts of the IDP inflows on household characteristics. There are six different dependent variables: access to drinkable water (column (1)), access to a toilet (column (2)), floor quality (column (3)), number of children under five (column (4)), households member (column (5)), and head of the household working (column (6)). The dependent variables are a binary, coded to one if the household has a certain characteristic. I use individual-level data on household characteristics and the date of the interview from the Demographic and Health Survey (DHS) from 1998 to 2017. Households interviewed from January 2008 are exposed to the treatment. The baseline specification is presented in equation (3), where I rely on within-district household variation. Panel A present the baseline results. Panel B shows the estimates with a continuous treatment: the predicted IDP inflows. Panel C adds an interaction if a child is IDP. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

for the needs of the local population (Din 2010). As a result, the arrival of new populations (IDP families) may have restricted access to quality health services for native and IDP populations, affecting the incidence of polio.

Conceptually, by capturing the aggregate demand in health services, I would be able to shed some light on the potential changes created by IDP inflows. Unfortunately, this information is not available. I can only capture the individual demand by using individual-level data on prenatal and postnatal doctor assistance from the DHS. Figure A.3.17 and Figure A.3.18 illustrate an increase in the share of children with prenatal assistance and a slight decrease in postnatal assistance after 2007 in host districts. However, I observe a decrease in prenatal and a more substantial decrease in postnatal services in non-host communities. To complement the descriptive analysis, I repeat equation (2) with the prenatal and postnatal doctor assistance outcomes (columns (1) and (4)). I include province-fixed effects in columns (2), (3), (5), and (6). Results in Table 3.5 (panel A) do not support the idea of changes in the individual demand for health services in treated districts after the IDP inflow. This result is unsurprising since it is hard to imagine that the IDP inflow generates a behavioural change in the demand for health services.

Is the supply responsive to an increase in the demand for health services? Ideally, I would like to study this question using district-level data on health service delivery (health centres and workforce). Unfortunately, I could not get this data. To address this limitation, I proxy health services supply with district-level data on polio vaccination campaigns from the Polio Eradication Program. In Pakistan, health

Table 3.5: Effect of IDP inflow on the demand and supply of health services

	(1)	(2)	(3)	(4)	(5)	(6)
	PANEL A: Demand health services					
VARIABLES	doctor prenatal	d. prenatal	d. prenatal	doctor postnatal	d. postnatal	d. postnatal
2007 x Host district	-0.037			0.032		
	(0.056)			(0.057)		
Predicted Inflow		0.003	0.003		0.014	0.013
		(0.014)	(0.014)		(0.017)	(0.017)
Predicted Inflow X IDP			0.007			0.019**
			(0.008)			(0.006)
Observations	10,623	10,623	10,623	10,623	10,623	10,623
District FE	Yes	No	No	Yes	No	No
Province FE	No	Yes	Yes	No	Yes	Yes
Controls	No	No	No	No	No	No
Number of districts	31	31	31	31	31	31
Number of provinces	7	7	7	7	7	7
	PANEL B: Number of polio cases					
VARIABLES	polio act.	polio act.	polio act.	polio act.	polio act.	
Predicted Inflow	0.112559***	0.066885***	0.060775***	0.065066***	0.062826***	
	(0.011082)	(0.004376)	(0.004399)	(0.009431)	(0.009756)	
Observations	8,184	8,184	8,184	6,600	5,040	
Province FE	No	Yes	Yes	Yes	Yes	
Time FE	No	Yes	Yes	Yes	Yes	
Controls	No	No	Yes	Yes	Yes	

Note: This Table presents the impacts of the IDP inflows on the demand and supply of health services. In Panel A I estimate equation (2) on individual-level data on prenatal (columns (1), (2), and (3)) and postnatal doctor assistance (columns (4), (5), and (6)). I exploit the within district cohort variation to identify the effects. I include province-fixed effects in columns (2), (3), (5), and (6). Panel B shows the estimates on the predicted IDP inflows on district-level data on polio vaccination campaigns from the Polio Eradication Program. The baseline specification is presented in equation (1) Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

is primarily the responsibility of the provincial government. Therefore, the central assumption is that the supply of health services follows the same pattern as the polio vaccination campaigns. Figure A.3.14 shows how the total vaccination activities increased in 2008 in host districts with respect 2007. I observe a similar increase in the total vaccination activities per 100.000 inhabitants (see Figure A.3.15).[13] Results of Table 3.5 support the idea of a responsive supply—an increase of one standard deviation in predicted inflows results in 0.06 additional vaccination campaign. However, I can not disentangle if the increase in the supply is high enough to meet the demand for formal health services.

3.6.4 Other mechanisms

After the conflict in F.A.T.A., the resistance to foreign interventions has considerably increased across the country. Along these lines, the misinformation and suspicion regarding the polio vaccination have also been a barrier to stopping polio eradication in Pakistan. These concerns include misconceptions regarding the vaccine's efficacy in local communities and among vaccinators engaged in repeat polio campaigns. Among the rumours, there is a widespread belief in Pakistan and elsewhere that the vaccine causes girls' infertility, impacting the reduction of vaccinated girls (Rahim, Ahmad, and Abdul-Ghafar 2022). The disclosure of the C.I.A. vaccination ruse could have significantly impacted mistrust of vaccines and "Western" health services because it lent credibility to many of the Taliban's arguments against

[13] I use the aggregate population stock in host and non-host districts using the 2017 Population Census. When I use the aggregate population using the 1998 Population Census, I obtain a similar graph in Figure A.3.16.

vaccines.[14] The mistrust in health services could be a potential underlying channel behind the increase in polio cases. Vaccine refusal has been mostly restricted to Khyber Pakhtunkhwa and F.A.T.A. Therefore, the IDP population from F.A.T.A. could transmit the misinformation to host communities. I repeat equation (1) to evaluate this hypothesis and restrict my time horizon until June 2011. Table A.3.11 present significant estimates when restricting my time horizon. Additionally, the results of Table 3.3 on immunization do not change when I drop the months after June 2011 (see panel B of Table A.3.11). Due to the infertility rumours, girls were the most affected by the vaccine refusal. Panel C of Table A.3.11 shows no statistical differences between host and no-host districts in the vaccination rates of girls. These findings help rule out the hypothesis that the IDPs could be active in misinformation transmission.

Finally, larger IDP inflows may increase prices through higher local demand or displace informal workers from the labour markets, thus deteriorating health outcomes through worse economic conditions. According to the Internal Monitoring Center, IDPs also face hostilities from the host communities. The perception exists that IDP workers substituted native workers because they accept lower pay than the local population and force wages down. Table 3.4 shows how the head of households in IDP families is less likely to have jobs than natives. But, this information does not allow me to examine the local labour market dynamics. So, given data constraints, I can not test this mechanism empirically.

3.7 Robustness Checks

I present evidence of the validity of my results in three ways. First, I conduct two falsification tests to rule out hidden effects. Second, I test the validity of my treatment and counterfactual using alternative definitions. Third, I approach my research question using alternative specifications.

3.7.1 Falsification Tests

In this project, I look at the impacts of hosting conflict-induced IDPs on polio incidence in host communities. A major concern is that the effects of Table 1 could be driven by the effect of conflict rather than by the IDP inflow. We should observe no effect on host districts before the treatment to reject this hypothesis. In this setting, the violence surged after the terrorist attack in September 2001. However,

[14] The C.I.A. wanted to obtain definite proof that Bin Laden was hiding in Abbottabad, Pakistan. To this end, the C.I.A. organised a vaccination ruse. The objective was to obtain D.N.A. samples of children living in the compound and compare them to the D.N.A. of Bin Laden's sister, who had died in Boston in 2010. On July 11th of 2011, the British newspaper *The Guardian* published an article describing the vaccine ruse (Martinez-Bravo and Stegmann 2022).

the mass movement of the population happened seven years later, after a sudden increase in the military offensive in 2008. I use the lag period between the beginning of the conflict (September 2001) and the onset of the IDP crisis (2008) to isolate the potential impact of the conflict. There is no effect on the number of polio cases before 2008 (see panel A of Table A.3.12).[15]

The peculiarity of this paper's setting is that most IDPs moved to districts within historical *Pashtunistan*. Suppose the large IDP inflows create the main effects. In that case, we should not observe an effect when comparing my counterfactual (red polygons in Figure A.3.17) to other non-Pashtu districts non-included in my baseline sample (white polygons in Figure A.3.17). The results of panel B of Table A.3.13 align with this assumption.

3.7.2 Treatment and Counterfactual Definition

The definition of my treatment relies on the historical border of *Pashtunistan*. As highlighted in section 4.1, I remove from my baseline sample the districts where only a share of their territory falls within *Pashtunistan*. These districts correspond to the red-dashed polygons in Figure A.3.17. I add these districts to my treatment definition. I show in Panel A of Table A.3.14 that although the magnitude of the effects decreases, the points estimates are statistically significant at the five per cent level. To facilitate a quick return and due to cultural barriers, IDP families settled in districts near F.A.T.A. Still, a minor share of IDPs moved to further districts within the historical *Pashtunistan*. The results of Table A.3.15 support this fact by presenting a very small or no effect when restricting the treatment to districts overlapping the Pashtu line.

I conduct two similar exercises to validate my counterfactual definition. First, I use the overlapping district to the Pashtu line as an alternative control group. The findings Table A.3.15 would imply that we should observe an adverse and significant effect when using this alternative counterfactual. Table A.3.16 test and validate this hypothesis (Panel C). Using the alternative counterfactual allows me to control for potential unobservables related to the Pashtu traditions and culture. Second, I show that the results hold when using the non-Pashtu districts not included in my baseline counterfactual as an alternative control group. See Table A.3.17 for further details.

3.7.3 Alternative Specification and Potential Cofounders

Polio cases have changed in the country over time, where the health response is the provincial government's responsibility. Thus, in my baseline specification (equation

[15] The results do not change when I use November 2004 as my falsification treatment timing.

1), I control for a year-month fixed effect to account for seasonal shocks standard across all districts in Pakistan and a provincial fixed effect to control for time-invariant characteristics within a district. However, my results are robust to alternative specifications. When controlling for the division fixed effect, only year fixed effect, or only province fixed effect, the magnitude of the effect is precisely the same. It only changes the significance level at the ten per cent level and increases the standard errors when I control for district-fixed effects (see panels A, B and C of Table A.3.18). Panel D shows how the standard errors decrease when I do not cluster them.

Finally, it could be a concern that IDP families chose their host community based on previous or existing polio cases number. Hence, a potential reverse causality could threaten my identification. Table A.3.19 rules out this hypothesis using the aggregate district polio cases from 2001 to 2007.

3.8 Conclusion

The increasing number of internally displaced people (IDP) poses new challenges to eradicating polio globally. This paper provides evidence that communities that received the IDP population increased their polio incidence, where the intensity of the inflow matters substantially.

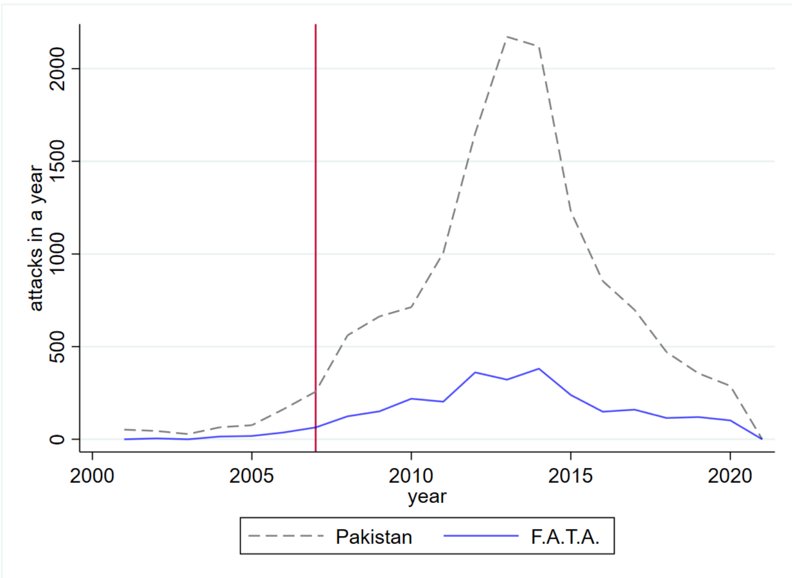
Three channels can explain these findings: First, although a vaccination programme targeting children on the move reduced the gaps in the immunization rates between IDP and native children, a sudden increase in the population in communities with low vaccination rates can facilitate the increase in polio incidence. Second, IDP families migrate to the poorest and most marginalized communities. The precarious health condition in the new destinations facilitates the spread of the virus. Third, the inflow of the IDP population could have congested health services in host communities.

The situation in Pakistan is not unique. In 2022, polio cases emerged in Malawi and Mozambique– two countries free of the virus for decades. Both countries are the scenario of a forced population movement from the conflicts ongoing in Northern Mozambique and the Eastern Democratic Republic of Congo, respectively. Protected children from diseases are far more likely to have the opportunity to thrive, the chance to learn and the ability to live healthy lives (UNICEF 2023). Therefore, three critical policy implications emerge from this paper. First, millions of forcibly displaced children migrate to camps or host communities. Since families in these settings are often transient, monitoring vaccination rates among these communities make it much harder to reach children with the necessary vaccines. Reaching the hard-to-reach- such as children from mobile and forced migrant populations or in conflict zones- should be a public priority (CDC 2021; UNICEF 2023). Second, poor communities are the host communities of most of the IDPs. An effort to better integrate the IDP population into the health services and labour market should be made to improve the conditions in which they live. Finally, the inflow of new population comes with increased demand for health services. Even if the increase in the demand is modest, in locations where the health delivery or capacity is weak, it can congest the local health services. It is essential to reinforce host communities' health workforce and infrastructure, so locals and newcomers can access health services equally.

Appendix Chapter 3

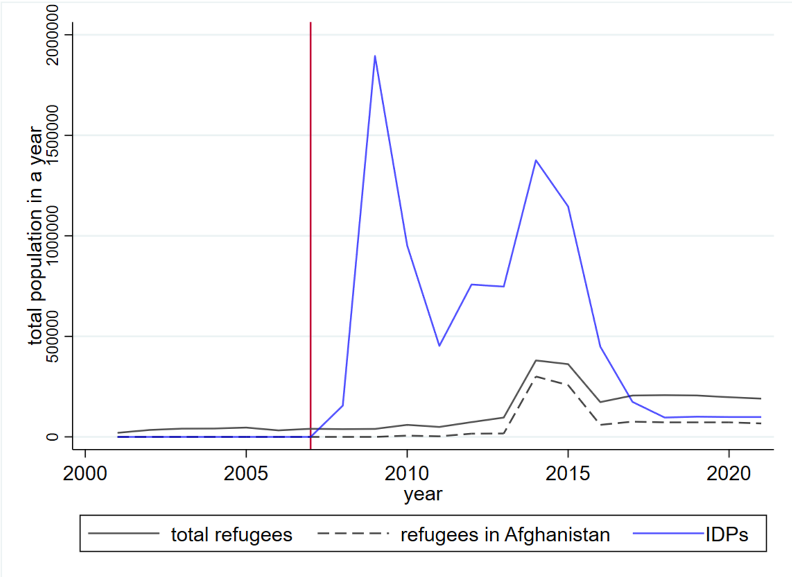
A Additional Figures

Figure A.3.1: Terrorist attacks (2001-2022)



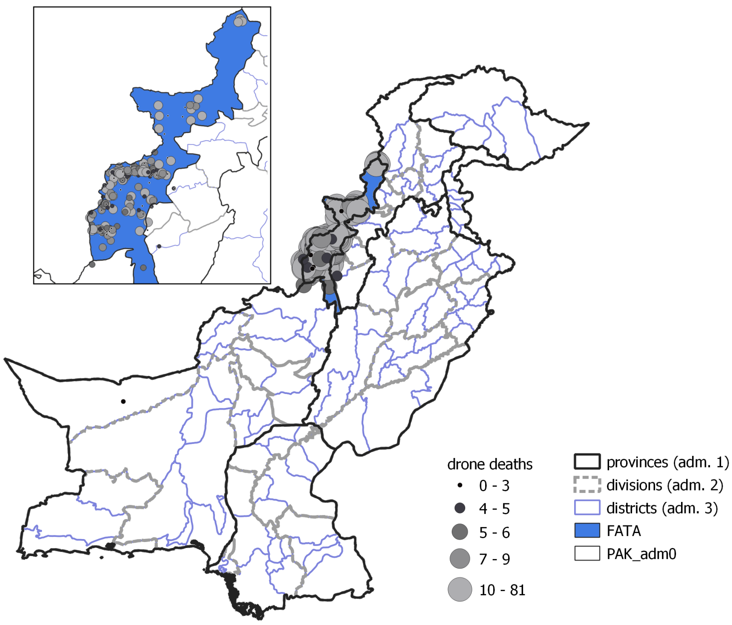
Note: This figure shows the yearly number of terrorist attacks from 2001 to 2022. The grey dashed line for Pakistan and the blue line for F.A.T.A. The vertical red line corresponds to 2007. Source. The Global Terrorism Database - G.T.D. (GTD 2021).

Figure A.3.2: Forcibly displaced population within and outside Pakistan (2001-2022)



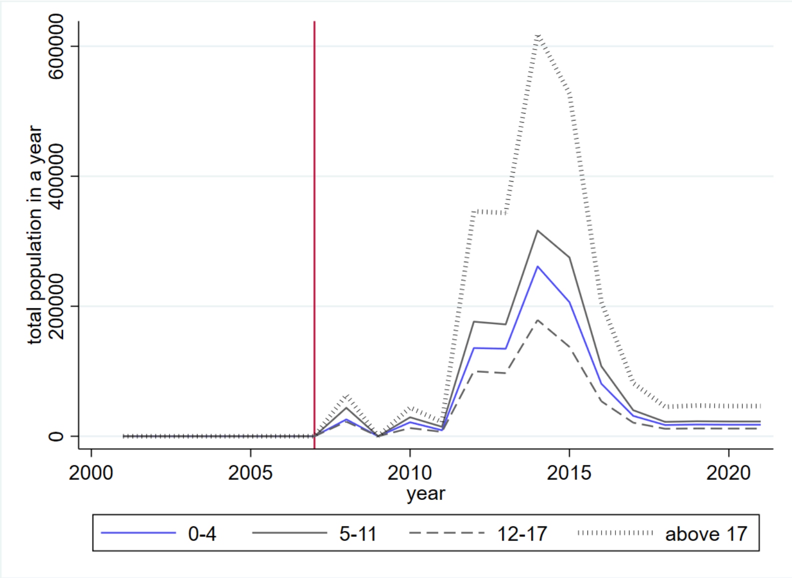
Note: This figure shows the yearly displaced population from Pakistan from 2001 to 2022. The blue line corresponds to the internally displaced persons (IDP). The black line shows the number of Pakistani refugees worldwide. And the black dashed line of the Pakistani refugees in Afghanistan. The vertical red line corresponds to 2007. Source. The United Nations High Commissioner for Refugees - U.N.H.C.R. (UNHCR 2022).

Figure A.3.3: Total deaths by drones (2001-2022)



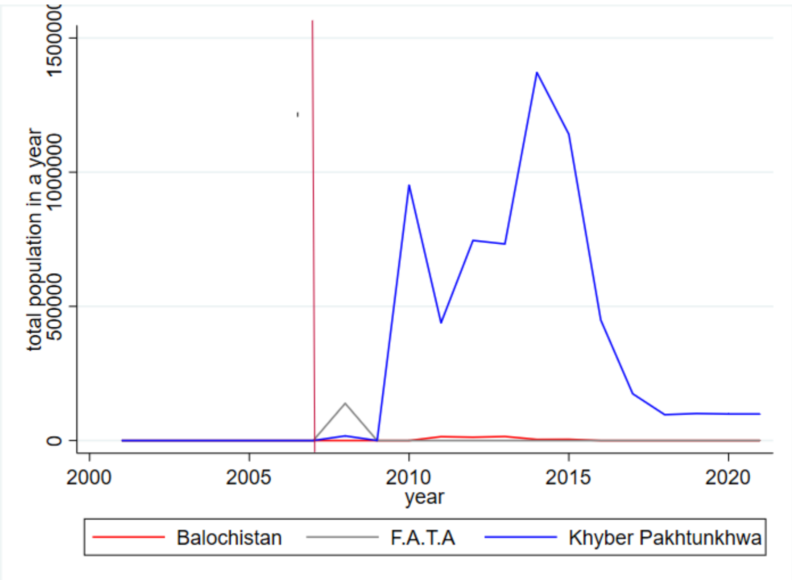
Note: This figure shows the spatial distribution of deaths associated with each drone strike from 2001 to 2022. The higher the dot higher is the total number of deaths. Source. The World of Drones Database developed by New America (New-America 2021).

Figure A.3.4: Internally Displaced Persons by age (2001-2022)



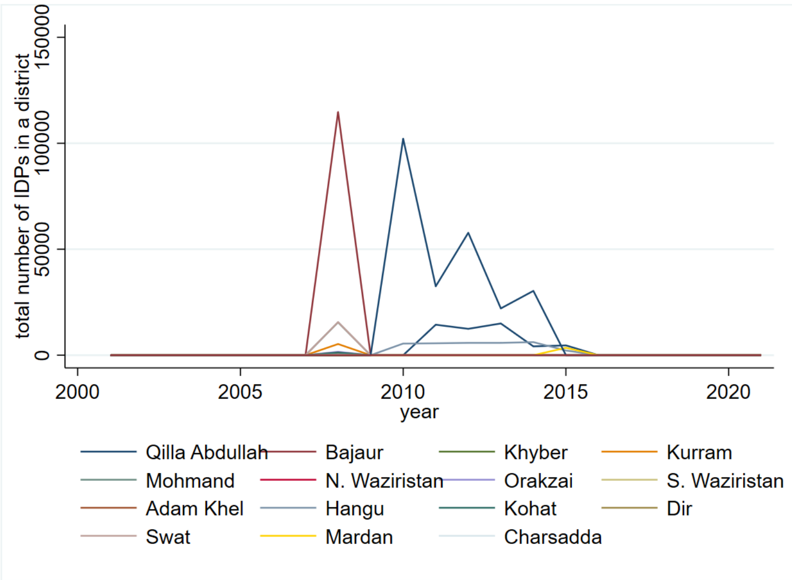
Note: This figure shows the yearly internally displaced population by age from 2001 to 2022. The blue line corresponds to the ages 0-4, the black line to the ages 5-11, the black dashed line to the ages 12-17 and the black pointed line to the ages above 17. The vertical red line corresponds to 2007. Source. The United Nations High Commissioner for Refugees - U.N.H.C.R. (UNHCR 2022).

Figure A.3.5: Internally Displaced Population by province



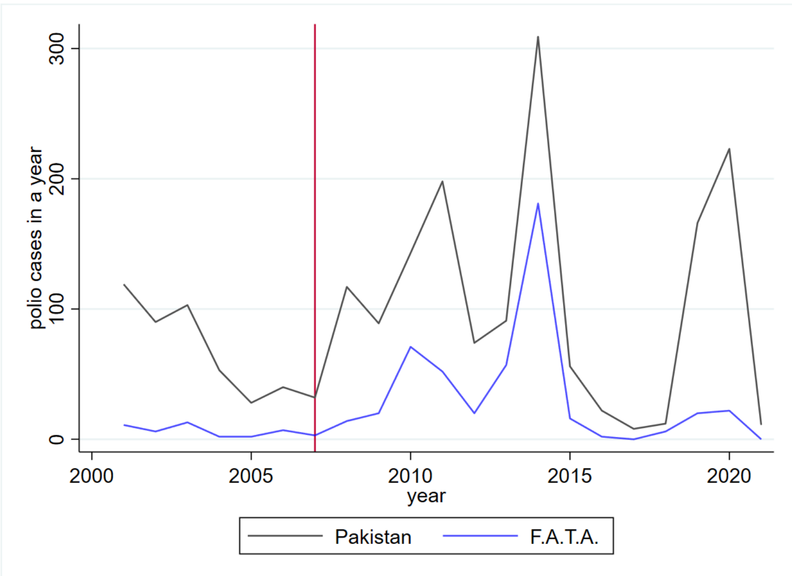
Note: This figure shows the yearly internally displaced population by province from 2001 to 2022. The red line corresponds to the province of Balochistan, border to Southern F.A.T.A.. The grey line are the IDPs in F.A.T.A. The blue line corresponds to Khyber Pakhtunkhwa province. Khyber Pakhtunkhwa is the border of the Eastern and Northern F.A.T.A. The vertical red line corresponds to 2007. Source. The United Nations High Commissioner for Refugees - U.N.H.C.R. (UNHCR 2022).

Figure A.3.6: Internally Displaced Population by destination district (2001-2022)



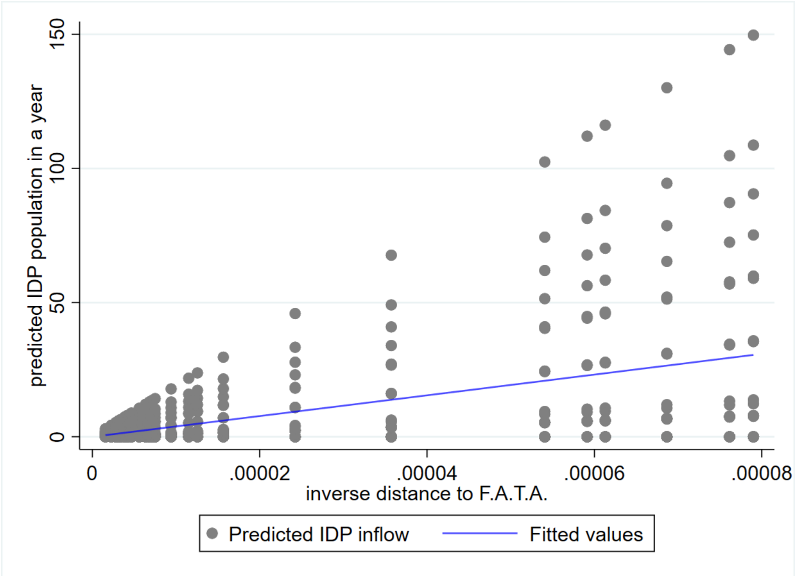
Note: This figure shows the yearly internally displaced population by district from 2001 to 2022. Source. The United Nations High Commissioner for Refugees - U.N.H.C.R. (UNHCR 2022).

Figure A.3.7: Polio cases in Pakistan and F.A.T.A. (2001-2022)



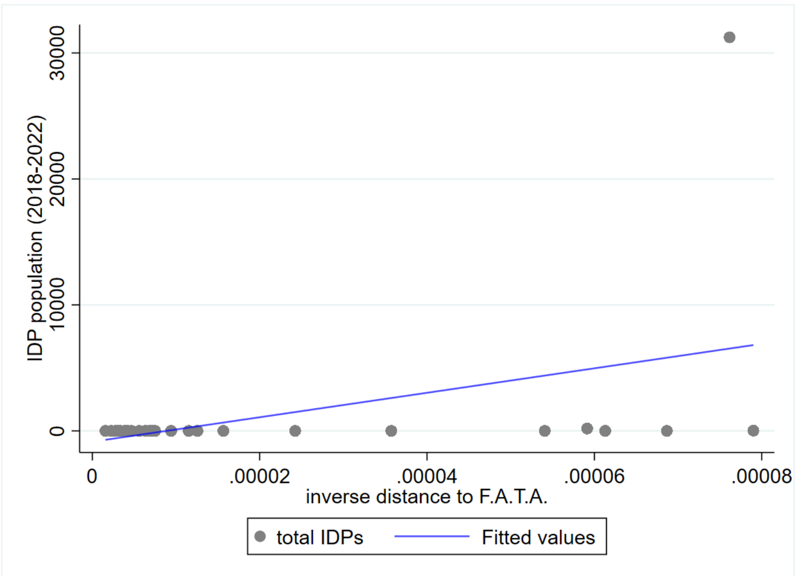
Note: This figure plots the yearly polio cases in Pakistan and F.A.T.A. Treated districts are the host districts, and control districts are the non-host districts. Source. The Polio Eradication Program established by the World Health Organization (WHO).

Figure A.3.8: Inverse distance and predicted inflow



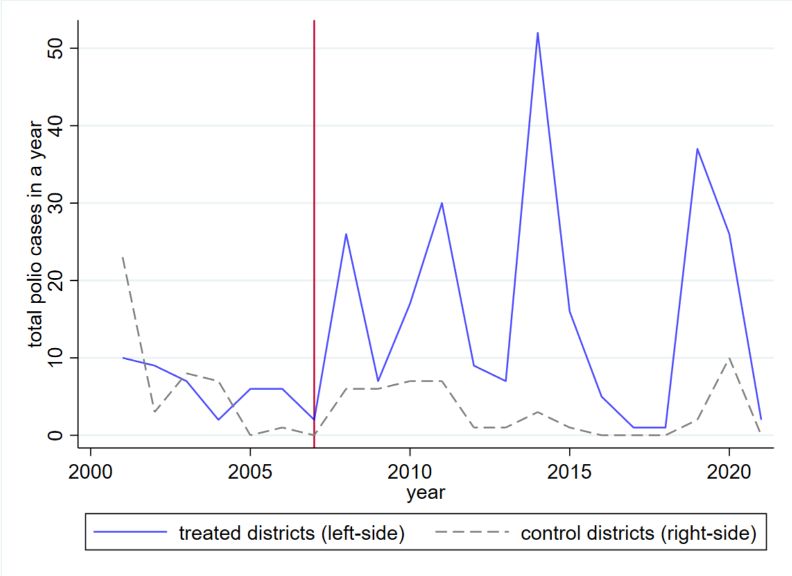
Note: This figure shows the correlation between the predicted inflow measure and the inverse distance to the closest F.A.T.A. border. The predicted inflow measure is equal to the interaction of the inverse distance of each district to the nearest F.A.T.A. border (district variation) and the total yearly number of IDP population (annual variation).

Figure A.3.9: Inversed distance and reported IDPs



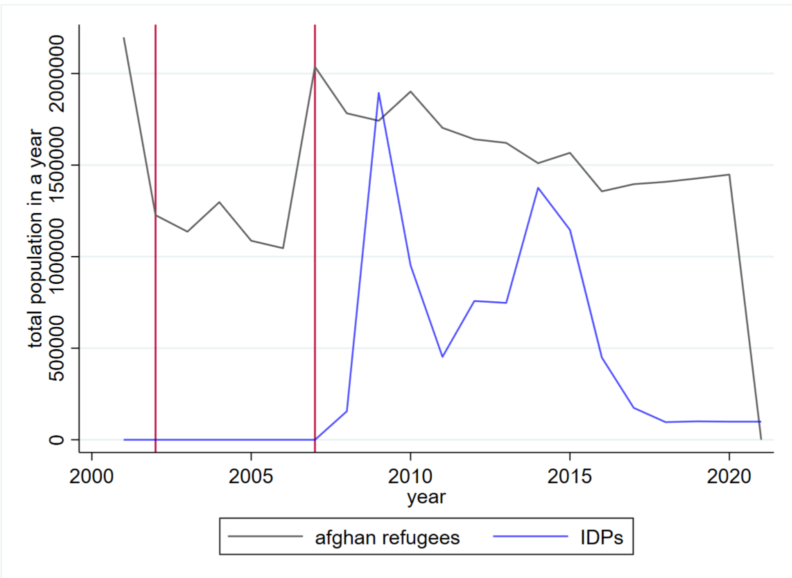
Note: This figure shows the correlation between the actual IDP inflow and the inverse distance to the closest F.A.T.A. border. The IDP information comes from the United Nations High Commissioner for Refugees - U.N.H.C.R. (UNHCR 2022).

Figure A.3.10: Polio cases (2001-2022)



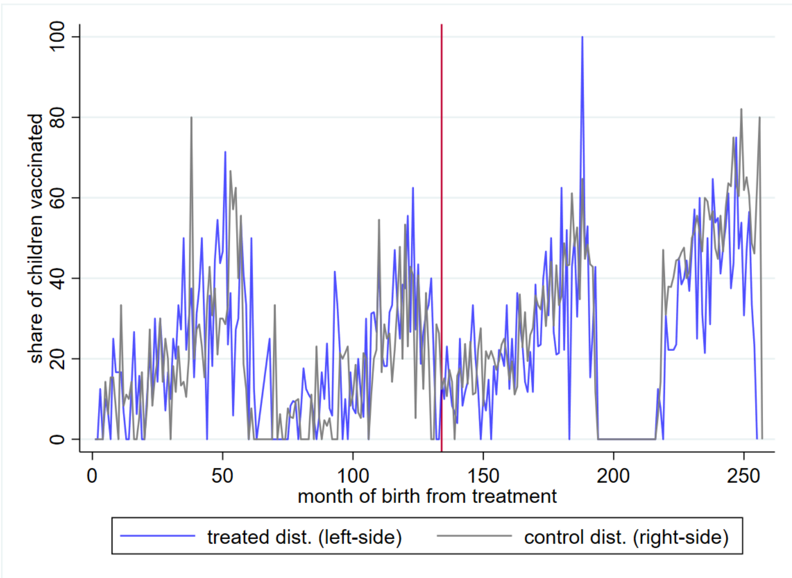
Note: This figure plots the yearly polio cases in treated and control districts. Treated districts are the host districts, and control districts are the non-host districts. Districts whose territory falls within the pre-colonial region of *Pashtunistan* are host districts. Non-host districts are those whose territory is outside *Pashtunistan* but adjacent to the historical border.

Figure A.3.11: Afghan refugees (2001-2022)



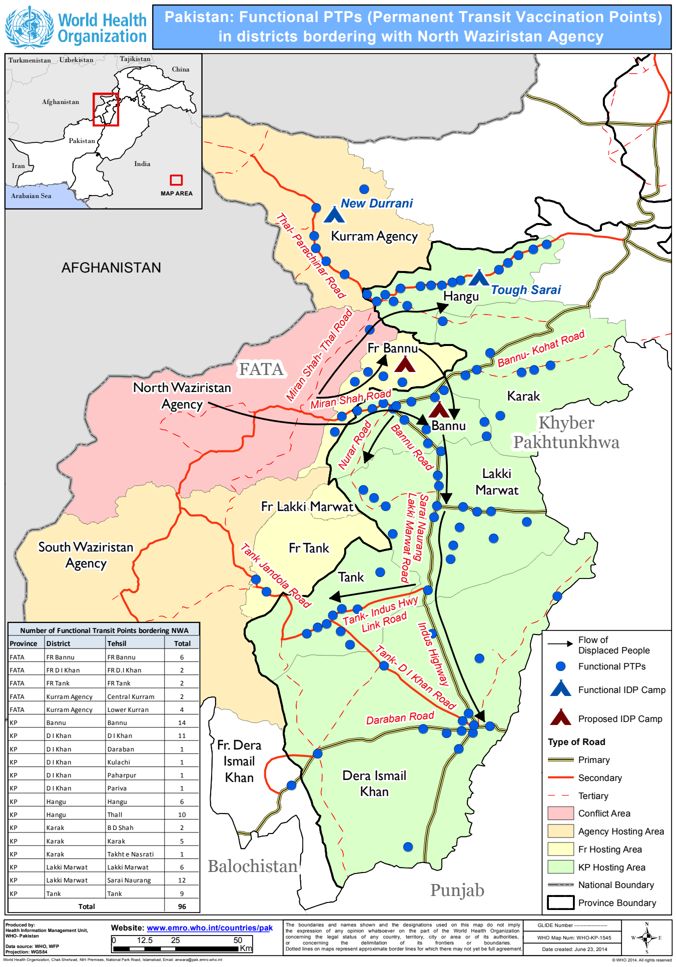
Note: This figure shows the yearly internally displaced population and Afghan refugees in Pakistan from 2001 to 2022. The blue line corresponds to the internally displaced persons (IDP). The black line indicates the number of Afghan refugees in Pakistan. The vertical red line corresponds to 2007. Source. The United Nations High Commissioner for Refugees - U.N.H.C.R. (UNHCR 2022).

Figure A.3.12: Inmunization rates (2001-2022), host vs non-host



Note: This figure plots the share of children vaccinated in treated and control districts by the birth cohort. Treated districts are the host districts and control districts are the non-host districts. Districts whose territory falls within the pre-colonial region of *Pashtunistan* are host districts. Non-host districts are those whose part is outside *Pashtunistan* but adjacent to the historical border. The vertical red line corresponds to December 2007. Source. Demographic and Health Survey (DHS).

Figure A.3.13: Permanent Transit Points (PTPs) to vaccinate children on the move

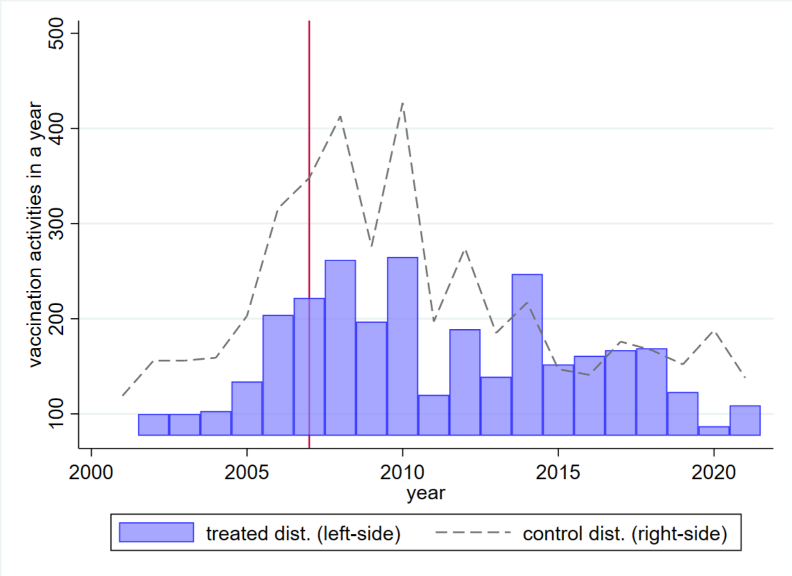


Number of Functional Transit Points bordering NWA

Province	District	Tehsil	Total
FATA	FR Bannu	FR Bannu	6
FATA	FR D I Khan	FR D.I Khan	2
FATA	FR Tank	FR Tank	2
FATA	Kurram Agency	Central Kurram	2
FATA	Kurram Agency	Lower Kurran	4
KP	Bannu	Bannu	14
KP	D I Khan	D I Khan	11
KP	D I Khan	Daraban	1
KP	D I Khan	Kulachi	1
KP	D I Khan	Paharpur	1
KP	D I Khan	Pariva	1
KP	Hangu	Hangu	6
KP	Hangu	Thall	10
KP	Karak	B D Shah	2
KP	Karak	Karak	5
KP	Karak	Takht e Nasrati	1
KP	Lakki Marwat	Lakki Marwat	6
KP	Lakki Marwat	Sarai Naurang	12
KP	Tank	Tank	9
Total			96

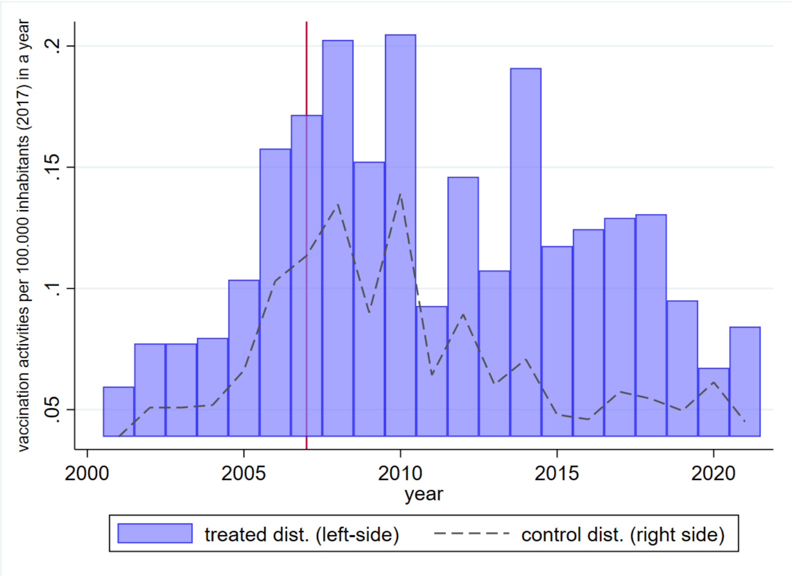
Note: This figure illustrates the functional Permanent Transit Vaccination Points in districts bordering the North Waziristan district. The Pakistan Polio Eradication Programme vaccinates children travelling or on the move through 500 permanent transit points (P.T.P.s) across all major transit points nationwide. These P.T.P.s are set up along country and district borders and other essential transit points such as railway stations, bus stops, and highways. Source. World Health Organization.

Figure A.3.14: Polio campaigns (2001-2022)



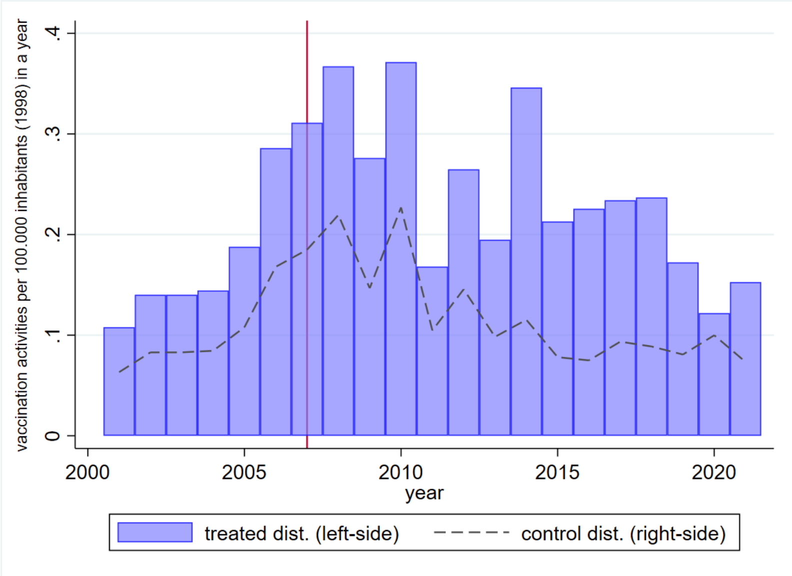
Note: This figure plots the total number of vaccination campaigns against polio in treated and control districts from 2001 to 2022. The blue bars show the campaigns in treated districts, and the grey dashed line in control districts. Treated districts are the host districts and control districts are the non-host districts. Districts whose territory falls within the pre-colonial region of *Pashtunistan* are host districts. Non-host districts are those whose part is outside *Pashtunistan* but adjacent to the historical border. The vertical red line corresponds to December 2007. Source. The Polio Eradication Program from the World Health Organization (WHO).

Figure A.3.15: Polio campaigns per 100,000 inhabitants (2001-2022)



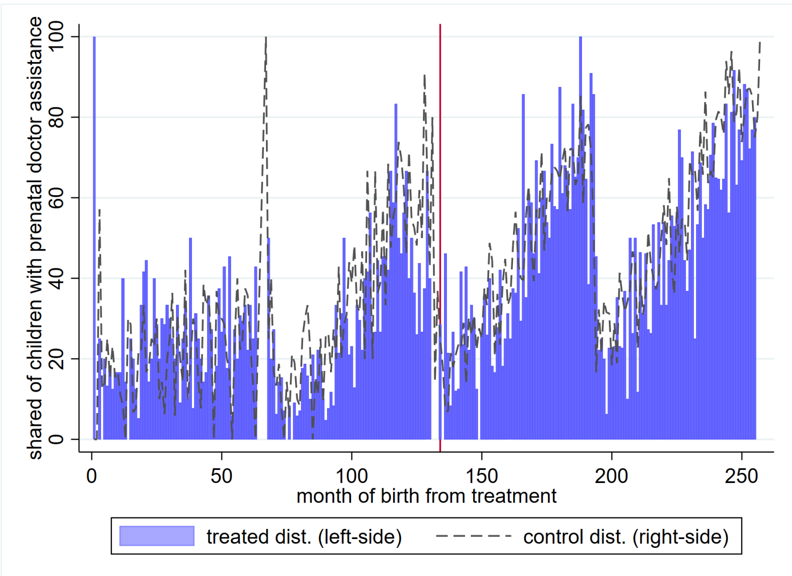
Note: This figure plots the number of vaccination campaigns against polio per 100,000 inhabitants in treated and control districts from 2001 to 2022. I calculate the campaigns per 100,000 inhabitants relative to the population in 2017 from the 2017 population census. The blue bars show the campaigns in treated districts, and the grey dashed line in control districts. Treated districts are the host districts and control districts are the non-host districts. Districts whose territory falls within the pre-colonial region of *Pashtunistan* are host districts. Non-host districts are those whose part is outside *Pashtunistan* but adjacent to the historical border. The vertical red line corresponds to December 2007. Source. The Polio Eradication Program from the World Health Organization (WHO).

Figure A.3.16: Polio campaigns per 100,000 inhabitants in 1998 (2001-2022)



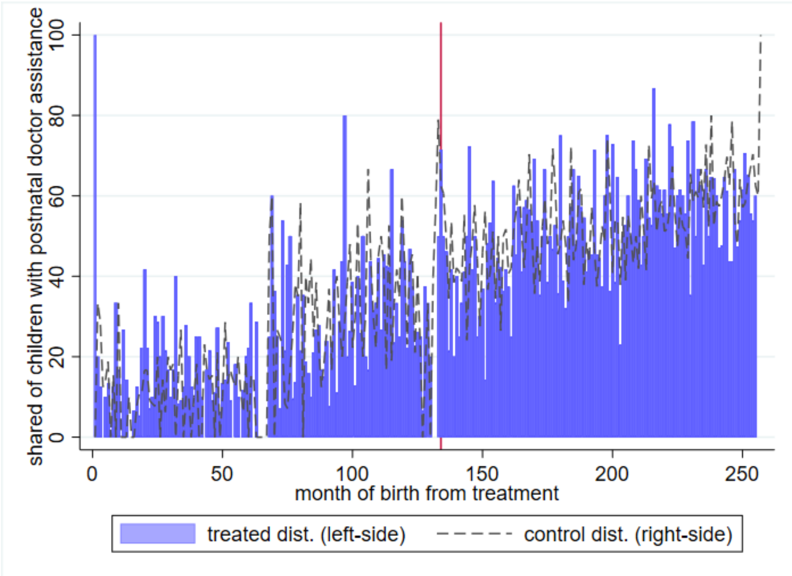
Note: This figure plots the number of vaccination campaigns against polio per 100,000 inhabitants in treated and control districts from 2001 to 2022. I calculate the campaigns per 100,000 inhabitants relative to the population in 1998 from the 1998 population census. The blue bars show the campaigns in treated districts, and the grey dashed line in control districts. Treated districts are the host districts and control districts are the non-host districts. Districts whose territory falls within the pre-colonial region of *Pashtunistan* are host districts. Non-host districts are those whose part is outside *Pashtunistan* but adjacent to the historical border. The vertical red line corresponds to December 2007. Source. The Polio Eradication Program from the World Health Organization (WHO).

Figure A.3.17: Shared of children with prenatal assistance



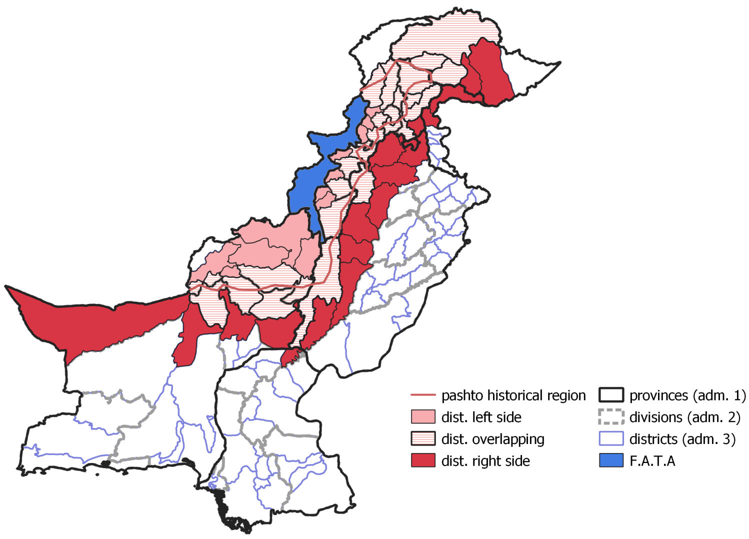
Note: This figure plots the share of children with prenatal assistance in treated and control districts by the birth cohort. Treated districts are the host districts and control districts are the non-host districts. Districts whose territory falls within the pre-colonial region of *Pashtunistan* are host districts. Non-host districts are those whose part is outside *Pashtunistan* but adjacent to the historical border. The vertical red line corresponds to December 2007. Source. Demographic and Health Survey (DHS).

Figure A.3.18: Shared of children with postnatal assistance



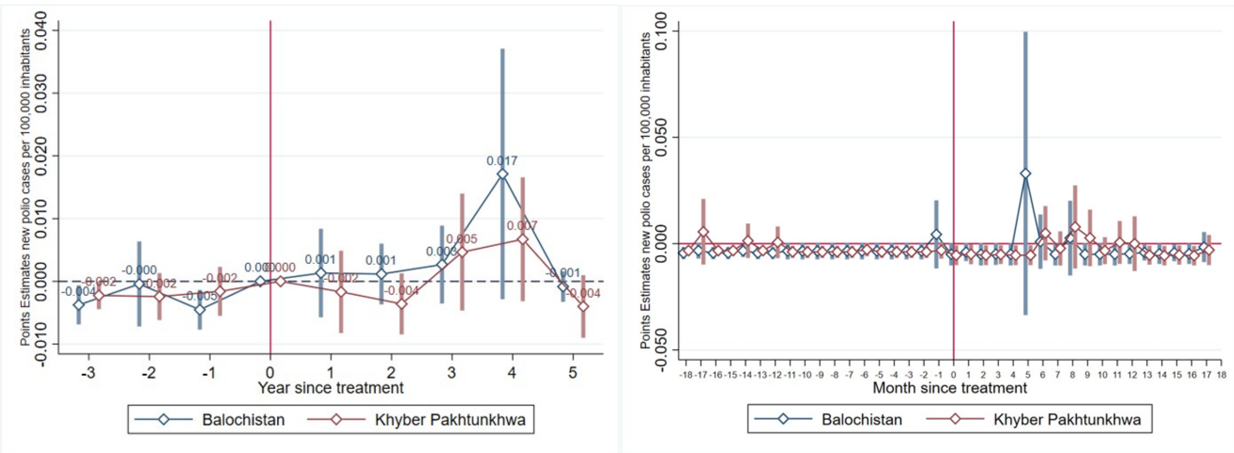
Note: This figure plots the share of children with postnatal assistance in treated and control districts by the birth cohort. Treated districts are the host districts and control districts are the non-host districts. Districts whose territory falls within the pre-colonial region of *Pashtunistan* are host districts. Non-host districts are those whose part is outside *Pashtunistan* but adjacent to the historical border. The vertical red line corresponds to December 2007. Source. Demographic and Health Survey (DHS).

Figure A.3.19: Alternative treated and control districts



Note: This figure shows the districts partially within *Pashtunistan*. To define them, I use the spatial distribution of districts relative to the pre-colonial region of *Pashtunistan*. The red line corresponds to the *Pashtunistan*'s border. The districts overlapping the border are the red dashed polygons, with only a share of the territory is in *Pashtunistan*. Districts whose territory falls within the pre-colonial region of *Pashtunistan* are host districts. Non-host districts are those whose territory is outside *Pashtunistan* but adjacent to the historical border.

Figure 3.25: Parallel trend test: Event study by treated province



Note: Figure A.3.20 plots the event and year coefficient by treated province from estimating equation 1 using the new polio cases per inhabitant as the dependent variable. The confidence intervals are 95%. Polio outcomes come from the Polio Eradication Program established by the World Health Organization (WHO). The omitted category is T-0, the year 2007. The dataset is in a year-district panel format. Treatment is defined at the year level. On the right-side diagram of this Figure, I repeat the exercise by month from the treatment.

B Additional Tables

Table A.3.1: Aggregate IDPs by district of origin (2001-2022)

(1)	(2)	(3)	(4)	(5)	(6)	(7)
Province and Division	Position	District	IDP families	IDP individuals	Drones	Total deaths
F.A.T.A	Southern	North Waziristan	108,149	648,894	291	2003
F.A.T.A	Southern	South Waziristan	71,124	426,744	84	678
F.A.T.A	Southern	Largha Shirani	0	0	1	6
F.A.T.A	Northern	Bajaur	72,895	437,370	4	128
F.A.T.A		Khyber	91,689	550,134	6	61
F.A.T.A		Kurram	33,024	198,144	9	83
F.A.T.A		Mohmand	36,759	220,554	0	0
F.A.T.A		Orakzai	35,823	214,938	1	13
N.W.F.P.	Southern	Tank	2,256	13,536	1	5
TOTAL			451,719	2,710,314	396	2,971

Note: This Table shows the aggregate number of internally displaced persons (IDP) from 2001 to 2022 by district of origin. The IDP data source is UNHCR 2022. Columns (6) and (7) present the aggregate number of drones and the number of deaths created from 2001 to 2022 from New-America 2021.

Table A.3.2: Total IDPs by district of destination (2008-2015)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Province	District	2008	2010	2011	2012	2013	2014	2015
Khyber Pakhtunkhwa (NWFP)	Adam Khel	1168						
Khyber Pakhtunkhwa (NWFP)	Charsadda	187						
Khyber Pakhtunkhwa (NWFP)	Dir	190						
Khyber Pakhtunkhwa (NWFP)	Hangu	63	5500	5635	5821	5821	6173	2232
Khyber Pakhtunkhwa (NWFP)	Kohat	1237						
Khyber Pakhtunkhwa (NWFP)	Mardan							3504
Khyber Pakhtunkhwa (NWFP)	Nowshera		102127	32499	57771	22076	30352	
Khyber Pakhtunkhwa (NWFP)	Peshawar	21						
Khyber Pakhtunkhwa (NWFP)	Swat	15639						
Khyber Pakhtunkhwa (FATA)	Bajaur	114717						
Khyber Pakhtunkhwa (FATA)	Khyber	110						
Khyber Pakhtunkhwa (FATA)	Kurram	5275						
Khyber Pakhtunkhwa (FATA)	Mohmand	15516						
Khyber Pakhtunkhwa (FATA)	N. Waziristan	11						
Khyber Pakhtunkhwa (FATA)	Orakzai	1632						
Khyber Pakhtunkhwa (FATA)	S. Waziristan	43						
Balochistan	Qilla Abdullah			14397	12438	14978	4166	4632
TOTAL		155809	107627	52531	76030	42875	40691	10368

Note: This Table shows the total number of internally displaced persons (IDP) from 2008 to 2015 by district of destination. The IDP data source is UNHCR 2022. There are no data for 2009 and after 2015.

Table A.3.3: Number of polio cases per 100,000 inhabitants in 1998

	(1)	(2)	(3)	(4)	(5)
VARIABLES	polio pop.den.	polio pop.den.	polio pop.den.	polio pop.den.	polio pop.den.
2007 x Host district	0.011547***	0.011547**	0.011469*	0.012928*	0.015614**
	(0.003284)	(0.005432)	(0.005983)	(0.006702)	(0.007465)
Observations	7,128	7,128	7,128	6,600	5,760
Province FE	No	Yes	Yes	Yes	Yes
Time FE	No	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes
Number of districts		27	27	25	27

Note: This Table presents the impacts of the IDP inflows on district polio cases per 100,000 inhabitants measured in 1998 in host districts compared to non-host district. I measure the total cases per 100,000 inhabitant, I use the total population at district level from the 1998 Population Census. Robust standard errors are in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A.3.4: Pre-treatment characteristics, host vs non-host districts

	(1) Non-Host	(2) Host	(3) Diff
monthly polio cases	0.026	0.042	0.019
	(0.261)	(0.219)	(0.011)
monthly number of polio campaigns	0.702	0.702	0.000
	(0.457)	(0.457)	(0.000)
nightlight intensity	6.939	7.831	0.495
	(5.437)	(5.841)	(1.060)
monthly number of drones	0.000	0.000	0.000
	(0.457)	(0.457)	(0.000)
monthly number of terrorist attacks	0.070	0.078	0.008
	(0.457)	(0.457)	(0.000)
Household Characteristics in 1998 - Census			
electricity	0.529	0.838	0.278
	(0.018)	(0.126)	(0.000)
piped water	0.152	0.308	-0.003
	(0.091)	(0.078)	(0.000)
own house	0.892	0.807	-0.040
	(0.026)	(0.086)	(0.000)
family size	9.598	11.540	3.317
	(0.795)	(0.871)	(0.000)
number children under 5	0.301	0.301	0.050
	(0.028)	(0.017)	(0.000)
head of hh literate	0.257	0.270	-0.068
	(0.047)	(0.038)	(0.000)
muslim	0.994	0.993	-0.003
	(0.001)	(0.003)	(0.000)
Pashtu mother tongue	0.056	0.816	0.622
	(0.080)	(0.207)	(0.000)
Household Characteristics in 2006 -DHS			
piped water	0.346	0.636	0.104*
	(0.279)	(0.214)	(0.044)
floor	0.379	0.389	0.085
	(0.284)	(0.161)	(0.163)
television	0.310	0.400	0.150***
	(0.254)	(0.152)	(0.018)
watch tv every week	0.206	0.318	0.209***
	(0.208)	(0.095)	(0.031)
radio	0.393	0.555	0.161
	(0.167)	(0.207)	(0.114)
head hh working	0.204	0.127	-0.048**
	(0.136)	(0.115)	(0.015)
number children under 5	2.258	3.032	0.764***
	(0.469)	(0.595)	(0.047)
number members	8.323	12.076	4.068***
	(1.435)	(2.163)	(0.724)
urban	0.288	0.510	0.265*
	(0.215)	(0.306)	(0.105)
Observations	1,596	1,008	2,604

Note: This table reports descriptive statistics for the main variables and sample considered in the baseline analysis. The analysis covers 31 districts from 2001 to 2022 at the monthly level (264 observations per district). For the pre-treatment analysis, I restrict my timeframe from 2001 to 2007. Pre-treatment characteristics are from the 1998 Population Census and the 2006-2007 Demographic and Health Survey (DHS).

Table A.3.5: Potential conflict confounding effect

	(1)	(2)	(3)	(4)	(5)
VARIABLES	polio	polio	polio	polio	polio
	PANEL A: Northern sample				
2007 x Host district	0.048***	0.048*	0.059	0.068	0.079
	(0.017)	(0.027)	(0.040)	(0.060)	(0.051)
Observations	3,432	3,432	3,432	2,112	2,388
Number of districts		13	13	8	13
	PANEL B:Southern sample				
2007 x Host district	0.036***	0.036*	0.036*	0.041*	0.015
	(0.011)	(0.017)	(0.018)	(0.020)	(0.025)
Observations	4,752	4,752	4,752	4,488	3,012
Number of districts		18	18	17	18
	PANEL C: Terrorist attacks controls				
2007 x Host district	0.025***	0.030**	0.035**	0.044**	0.043**
	(0.009)	(0.011)	(0.014)	(0.018)	(0.017)
Observations	8,184	8,184	8,184	6,600	6,060
Number of districts		31	31	25	31
Province FE	No	Yes	Yes	Yes	Yes
Time FE	No	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes

Note: This Table presents the results of Table 3.1 by controlling for conflict. Panel A shows the estimates restricting the sample to Northern district of my main sample. The districts in the North are: Abbottabad, Attok, Chakwal, Charsadda, Hangu, Islamabad, Kargil, Kupwara (Gilgit Wazarat), Malakand P.A., Muzaffarabad, Neelum, Peshawar, and Rawalpindi. Panel B shows the estimates for a sample of Southern districts (Bannu, Bhakkar, Bhittani, Bolan, Chagai, Dera Bugti, Kalat, Kashmore, Khushab, Layyah, Musakhel, Muzaffargarh, Pishin, Qilla Saifullah, Rajan Pur, Tank, Zhob, and Ziarat). Panel C controls for the number of terrorist attacks. Standard errors are clustered at the district level. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.3.6: Potential Afghan refugees effect

	(1)	(2)	(3)	(4)	(5)
VARIABLES	polio	polio	polio	polio	polio
	PANEL A: Control for total refugees in a district				
2007 x Host district	0.025***	0.036**	0.039**	0.048**	0.052**
	(0.009)	(0.014)	(0.017)	(0.021)	(0.023)
	PANEL B: Number of refugee camps fixed effects				
2007 x Host district	0.038***	0.038**	0.041**	0.050**	0.054**
	(0.009)	(0.014)	(0.017)	(0.021)	(0.023)
	PANEL C: Number of refugee camps interaction				
2007 x Host district	0.014	0.026*	0.028*	0.037*	0.030
	(0.009)	(0.013)	(0.015)	(0.018)	(0.025)
2007 x Host district X n. camps	0.006***	0.003	0.003	0.003	0.005
	(0.001)	(0.003)	(0.003)	(0.003)	(0.006)
Observations	8,184	8,184	8,184	6,600	6,060
Province FE	No	Yes	Yes	Yes	Yes
Time FE	No	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes
Number of districts		31	31	25	31

Note: This Table presents the results of Table 3.1 by controlling for the presence of refugees from Afghanistan. Panel A shows the estimates controlling for the number of refugees in a year. Panel B includes the number of camps fixed effects. In Panel C, I add an interaction to the number of refugee camps. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.3.7: Potential international migration effects

	(1)	(2)	(3)	(4)	(5)
VARIABLES	polio	polio	polio	polio	polio
2007 x Host district	0.038***	0.038**	0.041**	0.050**	0.054**
	(0.009)	(0.014)	(0.017)	(0.021)	(0.023)
Observations	8,184	8,184	8,184	6,600	6,060
Province FE	No	Yes	Yes	Yes	Yes
Time FE	No	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes
Number of districts		31	31	25	31

Note: This Table presents the results of Table 3.1 controlling for the number of Pakistani refugees. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.3.8: Effect of IDP population on vaccination against polio, with province fixed effect

	(1)	(2)	(3)	(4)	(5)
VARIABLES	vaccinated	vaccinated	vaccinated	vaccinated	vaccinated
	PANEL A: Dycotomic treatment				
2007 x Host district	-0.083***	-0.059	-0.031	-0.043	-0.020
	(0.017)	(0.056)	(0.059)	(0.051)	(0.052)
	PANEL B: Continuous treatment - Predicted inflow				
Predicted Inflow	0.021***	-0.004	-0.005	-0.012	-0.011
	(0.005)	(0.015)	(0.016)	(0.017)	(0.017)
Observations	10,608	10,608	10,608	10,563	10,563
Province FE	No	Yes	Yes	Yes	Yes
Time FE	No	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes
Number of provinces		7	7	7	7

Note: This Table presents the results of Table 3.3 by controlling for province fixed effect. Panel A shows the estimates with the baseline treatment. Panel B shows the results with the predicted inflow. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.3.9: Pre-treatment children characteristics at individual level, host vs non-host districts

	(1) Non-Host	(2) Host	(3) Diff
water piped	0.337	0.615	0.096
	(0.473)	(0.487)	(0.046)
toilet	0.339	0.456	0.248
	(0.473)	(0.498)	(0.158)
floor	0.413	0.390	-0.002
	(0.493)	(0.488)	(0.179)
television	0.359	0.485	0.214***
	(0.480)	(0.500)	(0.041)
watch tv every week	0.249	0.425	0.231***
	(0.432)	(0.495)	(0.011)
radio	0.390	0.488	0.153
	(0.488)	(0.500)	(0.143)
head hh working	0.208	0.072	-0.058*
	(0.406)	(0.258)	(0.025)
number children under 5	2.341	3.073	0.983***
	(1.442)	(2.045)	(0.179)
number members	8.636	11.309	4.316**
	(4.331)	(6.494)	(1.497)
mother education	0.448	0.374	0.002
	(0.846)	(0.785)	(0.089)
diarrhea	0.154	0.146	0.028
	(0.361)	(0.353)	(0.017)
fever	0.260	0.252	0.034
	(0.439)	(0.434)	(0.024)
head hh woman	0.066	0.033	-0.083
	(0.249)	(0.178)	(0.070)
urban	0.326	0.544	0.196***
	(0.469)	(0.498)	(0.031)
girl	0.490	0.471	-0.039**
	(0.500)	(0.499)	(0.014)
Observations	1,596	1,008	2,604

Note: This table reports descriptive statistics for the main variables and sample considered in the baseline analysis. The analysis covers 31 district from 2001 to 2022 at the monthly level (264 observations per district). For the pre-treatment analysis I restrict my timeframe from 2001 to 2007. Pre-treatment characteristics are from the 1991-1992 and 2006-2007 Demographic and Health Survey (DHS).

Table A.3.10: Effect of IDP population on polio

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	diarrhea	diarrhea	diarrhea	fever	fever	fever
	PANEL A: District fixed-effects					
2007 x Host district	-0.016			-0.148***		
	(0.033)			(0.036)		
Predicted Inflow		-0.020**	-0.020**		-0.024	-0.022
		(0.009)	(0.009)		(0.019)	(0.019)
Predicted Inflow x IDP			-0.027			-0.048***
			(0.020)			(0.016)
Number of districts	31	31	31	31	31	31
	PANEL B: Province fixed-effects					
2007 x Host district	0.016			-0.067**		
	(0.013)			(0.021)		
Predicted Inflow		-0.007	-0.006		-0.009	-0.008
		(0.006)	(0.006)		(0.007)	(0.006)
Predicted Inflow x IDP			-0.028***			-0.048***
			(0.006)			(0.005)
Number of provinces	7	7	7	7	7	7
Observations	10,623	10,623	10,623	10,623	10,623	10,623
Province FE	No	No	No	No	No	No
Time FE	No	No	No	No	No	No
Controls	No	No	No	No	No	No

Note: This Table presents the impacts of the IDP inflows on diarrhea and fever. I use individual level data from the Demographic health surveys. The outcome on diarrhea is equal to one if the child had diarrhea recently, zero otherwise. Fever is 1 if the child had fever the last two weeks. The baseline specification is presented in equation (2). Panel A shows the results with district fixed effects as in equation (2). In panel B, I control for province fixed effects. Standard errors are clustered at the province level. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.3.11: Mistrust on vaccines channel: cases per 100,000 inhabitants

	(1)	(2)	(3)	(4)	(5)
VARIABLES	polio	polio	polio	polio	polio
	PANEL A: Dycotomic treatment, 2001-2011				
2007 x Host district	0.005	0.005*	0.005*	0.005	0.005**
	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)
	PANEL B: Continuos treatment - Predicted Inflow, 2001-2011				
Predicted Inflow	0.000913*	0.000911***	0.000868***	0.000637*	0.000488
	(0.000511)	(0.000167)	(0.000179)	(0.000207)	(0.000258)
Observations	3,240	3,240	3,240	3,000	1,500
Province FE	No	Yes	Yes	Yes	Yes
Time FE	No	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes
Number of province		5	5	4	5
	PANEL C: Girls sample				
VARIABLES	vaccinated	vaccinated	vaccinated	vaccinated	vaccinated
Predicted Inflow	0.006	-0.004	-0.009	-0.004	-0.021
	(0.009)	(0.010)	(0.014)	(0.010)	(0.013)
Observations	3,064	3,064	10,608	3,046	3,046
Province FE	No	Yes	Yes	Yes	Yes
Time FE	No	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes
Number of districts		26	31	26	26

Note: This Table presents the potential effect of vaccine mistrust. In panel A, I repeat the estimates of panel C of Table 3.1 in a timeframe from 2001 to 2011. In panel B, I repeat the estimates of Panel C of Table 3.2 restricting my timeframework from 2001 to 2011. Panel C shows the results of of panel A of Table 3.3 from 2001 to 2011 in a sample of girls. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.3.12: Falsification test: Effects before treatment

	(1)	(2)	(3)	(4)	(5)
			PANEL A: At least one polio case		
VARIABLES	polio	polio	polio	polio	polio
2007 x Host district	-0.017	-0.017	-0.014	-0.008	-0.085
	(0.029)	(0.025)	(0.021)	(0.015)	(0.094)
Observations	2,604	2,604	2,604	2,100	1,620
Number of districts		31	31	25	25
			PANEL B: Number of polio cases		
VARIABLES	polio cases	polio cases	polio cases	polio cases	polio cases
2007 x Host district	-0.035	-0.035	-0.030	-0.028	-0.120
	(0.034)	(0.031)	(0.025)	(0.021)	(0.106)
Observations	2,604	2,604	2,604	2,100	1,620
Number of districts		31	31	25	25 height
	PANEL C: Number of polio cases per 100,000 inhabitants				
VARIABLES	polio pop.den.	polio pop.den.	polio pop.den.	polio pop.den.	polio pop.den.
2007 x Host district	-0.000	-0.000	0.001	0.001	-0.003
	(0.002)	(0.003)	(0.002)	(0.002)	(0.008)
Observations	2,268	2,268	2,268	2,100	1,620
Number of districts		27	27	25	25
Province FE	No	Yes	Yes	Yes	Yes
Time FE	No	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes

Note: This Table presents the impacts of the IDP inflows on district polio prevalence in host districts compared to non-host district with a placebo treatment timing. The treatment timing starts from September 2001. I use the spatial distribution of districts with respect the pre-colonial region of *Pashtunistan* to define host and non-host districts. Districts whose territory fall within the pre-colonial region of *Pashtunistan* are defined as host districts. Non-host districts are the district whose territory is outside *Pashtunistan*, but are adjacent to the historical border. Observations are at the district and month level from 2001 to 2022. The baseline specification is presented in equation (1). Column (1) presents the results without province, year-month fixed effects and covariates. Column (2) includes province and year-month fixed effects. Column (3) controls for nightlight intensity and total vaccination campaigns. Columns (4) controls for pre-treatment district-covariates (the average number of children under five, the average number of members in a household, and the total share of the literate population from 1973, 1981 and 1998 Population Census). Column (5) control instead for contemporary characteristics (the average number of children under five, the average number of members in a household, shared households with piped water, and shared households with a finished floor). This Table present three different outcomes: at least one case of polio (panel A), total number of polio cases (panel B) and polio cases per 100,000 inhabitants from the 2017 Population Census (panel C). Standard errors are clustered at the district level. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.3.13: Falsification test: non-pashtu districts counterfactual

	(1)	(2)	(3)	(4)	(5)
	PANEL A: At least one polio case				
VARIABLES	polio	polio	polio	polio	polio
2007 x Host district	-0.009*	-0.009	-0.006	-0.002	0.000
	(0.005)	(0.012)	(0.012)	(0.012)	(0.012)
Observations	26,928	26,928	26,928	21,384	19,200
Number of districts		102	102	81	100
	PANEL B: Number of polio cases				
VARIABLES	polio cases	polio cases	polio cases	polio cases	polio cases
2007 x Host district	-0.038***	-0.038	-0.027	-0.016	-0.013
	(0.009)	(0.032)	(0.025)	(0.019)	(0.018)
Observations	26,928	26,928	26,928	21,384	19,200
Number of districts		102	102	81	100
	PANEL C: Number of polio cases per 100,000 inhabitants				
VARIABLES	polio pop.den.	polio pop.den.	polio pop.den.	polio pop.den.	polio pop.den.
2007 x Host district	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	3,960	3,960	3,960	3,696	3,468
Number of districts		15	15	14	15
Province FE	No	Yes	Yes	Yes	Yes
Time FE	No	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes

Note: This Table presents the impacts of the IDP inflows on district polio prevalence in host districts compared to non-host district with a placebo counterfactual. The treatment timing starts from 2008. I use the spatial distribution of districts with respect the pre-colonial region of *Pashtunistan* to define host and non-host districts. Treated districts are the district whose territory is outside *Pashtunistan*, but are adjacent to the historical border. They correspond to the non-host districts in equation (1). Control districts are the district whose territory is outside *Pashtunistan*, but are not adjacent to the historical border. These districts are not included in my baseline sample. The baseline specification is presented in equation (1). Column (1) presents the results without province, year-month fixed effects and covariates. Column (2) includes province and year-month fixed effects. Column (3) controls for nightlight intensity and total vaccination campaigns. Columns (4) controls for pre-treatment district-covariates (the average number of children under five, the average number of members in a household, and the total share of the literate population from 1973, 1981 and 1998 Population Census). Column (5) control instead for contemporary characteristics (the average number of children under five, the average number of members in a household, shared households with piped water, and shared households with a finished floor). This Table present three different outcomes: at least one case of polio (panel A), total number of polio cases (panel B) and polio cases per 100,000 inhabitants from the 2017 Population Census (panel C). Standard errors are clustered at the district level. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.3.14: Treatment definition: districts entirely or partially in *Pashtunistan*

	(1)	(2)	(3)	(4)	(5)
	PANEL A: At least one polio case				
VARIABLES	polio	polio	polio	polio	polio
2007 x Host district	0.021**	0.022**	0.026*	0.022**	
	(0.005)	(0.006)	(0.007)	(0.009)	(0.008)
Observations	15,312	15,312	15,312	12,672	11,304
Number of districts		58	58	48	57
	PANEL B: Number of polio cases				
VARIABLES	polio cases	polio cases	polio cases	polio cases	polio cases
2007 x Host district	0.044***	0.044**	0.042**	0.051**	0.044***
	(0.010)	(0.013)	(0.012)	(0.016)	(0.012)
Observations	15,312	15,312	15,312	12,672	11,304
Number of districts		58	58	48	57
	PANEL C: Number of polio cases per 100,000 inhabitants				
VARIABLES	polio pop.den.	polio pop.den.	polio pop.den.	polio pop.den.	polio pop.den.
2007 x Host district	0.006575***	0.006575*	0.006661*	0.007462	0.009270**
	(0.001872)	(0.002769)	(0.003037)	(0.004243)	(0.003327)
Observations	7,128	7,128	7,128	6,600	5,760
Number of districts		27	27	25	27
Province FE	No	Yes	Yes	Yes	Yes
Time FE	No	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes

Note: This Table presents the impacts of the IDP inflows on district polio prevalence in an alternative defnition of host districts compared to baseline definition of non-host district. Host districts are the district whose territory is entirely or partially inside *Pashtunistan*. The treatment timing starts from 2008. Observations are at the district and month level from 2001 to 2022. The baseline specification is presented in equation (1). Column (1) presents the results without province, year-month fixed effects and covariates. Column (2) includes province and year-month fixed effects. Column (3) controls for nightlight intensity and total vaccination campaigns. Columns (4) controls for pre-treatment district-covariates (the average number of children under five, the average number of members in a household, and the total share of the literate population from 1973, 1981 and 1998 Population Census). Column (5) control instead for contemporary characteristics (the average number of children under five, the average number of members in a household, shared households with piped water, and shared households with a finished floor). This Table present three different outcomes: at least one case of polio (panel A), total number of polio cases (panel B) and polio cases per 100,000 inhabitants from the 2017 Population Census (panel C). Standard errors are clustered at the district level. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.3.15: Treatment definition: districts partially in *Pashtunistan*

	(1)	(2)	(3)	(4)	(5)
	PANEL A: At least one polio case				
VARIABLES	polio	polio	polio	polio	polio
2007 x Host district	0.013**	0.013**	-0.001	0.017	0.012
	(0.006)	(0.004)	(0.005)	(0.007)	(0.008)
Observations	12,144	12,144	15,312	9,768	9,012
Number of districts		46	58	37	45
	PANEL B: Number of polio cases				
VARIABLES	polio cases	polio cases	polio cases	polio cases	polio cases
2007 x Host district	0.027**	0.027**	-0.005	0.035**	0.025*
	(0.012)	(0.008)	(0.014)	(0.010)	(0.012)
Observations	12,144	12,144	15,312	9,768	9,012
Number of districts		46	58	37	45
	PANEL C: Number of polio cases per 100,000 inhabitants				
VARIABLES	polio pop.den.	polio pop.den.	polio pop.den.	polio pop.den.	polio pop.den.
2007 x Host district	-0.000163	-0.000000	0.001534	-0.000569	-0.000454
	(0.000942)	(0.000000)	(0.001007)	(0.001017)	(0.001830)
Observations	3,960	3,960	7,128	3,696	3,468
Number of districts		15	27	14	15
Province FE	No	Yes	Yes	Yes	Yes
Time FE	No	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes

Note: This Table presents the impacts of the IDP inflows on district polio prevalence in an alternative defnition of host districts compared to baseline definition of non-host district. Host districts are the district whose territory is partially inside *Pashtunistan*. The treatment timing starts from 2008. Observations are at the district and month level from 2001 to 2022. The baseline specification is presented in equation (1). Column (1) presents the results without province, year-month fixed effects and covariates. Column (2) includes province and year-month fixed effects. Column (3) controls for nightlight intensity and total vaccination campaigns. Columns (4) controls for pre-treatment district-covariates (the average number of children under five, the average number of members in a household, and the total share of the literate population from 1973, 1981 and 1998 Population Census). Column (5) control instead for contemporary characteristics (the average number of children under five, the average number of members in a household, shared households with piped water, and shared households with a finished floor). This Table present three different outcomes: at least one case of polio (panel A), total number of polio cases (panel B) and polio cases per 100,000 inhabitants from the 2017 Population Census (panel C). Standard errors are clustered at the district level. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.3.16: Counterfactual definition: districts partially in *Pashtunistan*

	(1)	(2)	(3)	(4)	(5)
			PANEL A: At least one polio case		
VARIABLES	polio	polio	polio	polio	polio
2007 x Host district	0.025***	0.025*	0.026	0.030	0.048
	(0.009)	(0.010)	(0.012)	(0.018)	(0.021)
Observations	10,296	10,296	10,296	8,976	7,536
Number of districts		39	39	34	38
			PANEL B: Number of polio cases		
VARIABLES	polio cases	polio cases	polio cases	polio cases	polio cases
2007 x Host district	0.054***	0.054	0.051	0.054	0.094
	(0.016)	(0.029)	(0.030)	(0.040)	(0.056)
Observations	10,296	10,296	10,296	8,976	7,536
Number of districts		39	39	34	38
			PANEL C: Number of polio cases per 100,000 inhabitants		
VARIABLES	polio pop.den.	polio pop.den.	polio pop.den.	polio pop.den.	polio pop.den.
2007 x Host district	0.005878***	0.005874***	0.005481*	0.006283	0.009007*
	(0.002034)	(0.002791)	(0.002548)	(0.006168)	(0.007225)
Observations	3,168	3,168	3,168	2,904	2,292
Number of districts		12	12	11	12
Province FE	No	Yes	Yes	Yes	Yes
Time FE	No	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes

Note: This Table presents the impacts of the IDP inflows on district polio prevalence in the baseline defnition of host districts compared to an alternative definition of non-host district. Non-host districts are the district whose territory is partially inside *Pashtunistan*. The treatment timing starts from 2008. Observations are at the district and month level from 2001 to 2022. The baseline specification is presented in equation (1). Column (1) presents the results without province, year-month fixed effects and covariates. Column (2) includes province and year-month fixed effects. Column (3) controls for nightlight intensity and total vaccination campaigns. Columns (4) controls for pre-treatment district-covariates (the average number of children under five, the average number of members in a household, and the total share of the literate population from 1973, 1981 and 1998 Population Census). Column (5) control instead for contemporary characteristics (the average number of children under five, the average number of members in a household, shared households with piped water, and shared households with a finished floor). This Table present three different outcomes: at least one case of polio (panel A), total number of polio cases (panel B) and polio cases per 100,000 inhabitants from the 2017 Population Census (panel C). Standard errors are clustered at the district level. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.3.17: Counterfactual definition: non-Pashtu districts not adjacent to *Pashtunistan* border

	(1)	(2)	(3)	(4)	(5)
			PANEL A: At least one polio case		
VARIABLES	polio	polio	polio	polio	polio
2007 x Host district	0.029***	0.029*	0.026*	0.038*	0.058***
	(0.009)	(0.014)	(0.013)	(0.015)	(0.014)
Observations	25,080	25,080	25,080	20,592	17,724
Number of districts		95	95	78	93
			PANEL B: Number of polio cases		
VARIABLES	polio cases	polio cases	polio cases	polio cases	polio cases
2007 x Host district	0.043***	0.043	0.033	0.071	0.115**
	(0.014)	(0.041)	(0.040)	(0.039)	(0.047)
Observations	25,080	25,080	25,080	20,592	17,724
Number of districts		95	95	78	93
			PANEL C: Number of polio cases per 100,000 inhabitants		
VARIABLES	polio pop.den.	polio pop.den.	polio pop.den.	polio pop.den.	polio pop.den.
2007 x Host district	0.004565**	0.004002*	0.005048*	0.006477*	0.009815*
	(0.001872)	(0.002769)	(0.002287)	(0.003541)	(0.003602)
Observations	3,168	3,168	3,168	2,904	2,292
Number of districts		12	12	11	12
Province FE	No	Yes	Yes	Yes	Yes
Time FE	No	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes

Note: This Table presents the impacts of the IDP inflows on district polio prevalence in the baseline defnition of host districts compared to an alternative definition of non-host district. Non-host districts are the district outside *Pashtunistan*, and non-adjacent to *Pashtunistan* border. The treatment timing starts from 2008. Observations are at the district and month level from 2001 to 2022. The baseline specification is presented in equation (1). Column (1) presents the results without province, year-month fixed effects and covariates. Column (2) includes province and year-month fixed effects. Column (3) controls for nightlight intensity and total vaccination campaigns. Columns (4) controls for pre-treatment district-covariates (the average number of children under five, the average number of members in a household, and the total share of the literate population from 1973, 1981 and 1998 Population Census). Column (5) control instead for contemporary characteristics (the average number of children under five, the average number of members in a household, shared households with piped water, and shared households with a finished floor). This Table present three different outcomes: at least one case of polio (panel A), total number of polio cases (panel B) and polio cases per 100,000 inhabitants from the 2017 Population Census (panel C). Standard errors are clustered at the district level. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.3.18: Alternative specifications

	(1)	(2)	(3)	(4)	(5)
VARIABLES	polio	polio	polio	polio	polio
	PANEL A: Fixed effect and Cluster at division level				
2007 x Host district	0.038***	0.038**	0.053*	0.060*	0.070*
	(0.009)	(0.017)	(0.028)	(0.031)	(0.037)
Observations	8,184	8,184	8,184	6,600	6,060
Number of division		19	19	15	19
Division FE	No	Yes	Yes	Yes	Yes
Time FE	No	Yes	Yes	Yes	Yes
	PANEL B: Province and Year fixed effects				
2007 x Host district	0.038***	0.038**	0.041**	0.049**	0.052**
	(0.009)	(0.014)	(0.017)	(0.020)	(0.022)
Observations	8,184	8,184	8,184	6,600	6,060
Number of districts		31	31	25	31
Province FE	No	Yes	Yes	Yes	Yes
Time FE	No	Yes	Yes	Yes	Yes
	PANEL C: Province fixed effects				
2007 x Host district	0.038***	0.038**	0.039**	0.047**	0.068**
	(0.009)	(0.014)	(0.016)	(0.019)	(0.026)
Observations	8,184	8,184	8,184	6,600	6,060
Number of districts		31	31	25	31
Province FE	No	Yes	Yes	Yes	Yes
	PANEL D: No clusters				
2007 x Host district	0.038***	0.038***	0.041***	0.050***	0.054***
	(0.009)	(0.008)	(0.008)	(0.010)	(0.013)
Observations	8,184	8,184	8,184	6,600	6,060
Number of districts		31	31	25	31
Province FE	No	Yes	Yes	Yes	Yes
Time FE	No	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes

Note: This Table presents the results of Table 3.1 for different specifications. Panel A shows the estimates controlling division fixed effect instead of province fixed effect. Panel B control only for province and year fixed effects. In Panel C, I only control for province fixed effect. Panel D shows the estimates without clustering the error terms. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.3.19: Potential reverse causality

	(1)	(2)	(3)	(4)	(5)
VARIABLES	predicted inflow	predicted inflow	predicted inflow	predicted inflow	predicted inflow
yearly polio cases	0.031668***	0.014447	0.011966	-0.007485	0.010872
	(0.003580)	(0.017856)	(0.016618)	(0.003425)	(0.016732)
Observations	8,184	8,184	8,184	6,600	5,040
Province FE	No	Yes	Yes	Yes	Yes
Time FE	No	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes
Number of provinces		7	7	4	7

Note: This Table presents the estimates of the yearly polio cases from 2001 to 2007 on the predicted inflow. This estimation allows me to evaluate the potential reverse causality of the number of cases on the decision on where to migrate. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Conclusion

This thesis shows that internal displacement adversely affects intra-household and intra-community behaviours, impacting not only internally displaced persons (IDP), but also their communities of origin and destination.

Throughout this thesis, I generate compelling evidence on how internal displacement can change intra-household decisions on when young women marry, as well as the long-term social cohesion within host communities. Additionally, this thesis provides suggestive evidence of how the sudden arrival of IDPs can disrupt the government's efforts to eradicate infectious diseases.

Chapter 1 provides evidence that exposing young women to events that generate large forcibly displaced population outflows- such as natural disasters, conflicts, or human-made disasters- leads to early marriage. The impacts hold for both women in the sending locations and internally displaced women. Interestingly, the effect on IDP women is even stronger compared to those who stay in the communities of origin. The displacement is the turning point for IDP women. Forced displacement negatively impacts IDP households. In turn, to cope with this shock, IDP women marry earlier to maximize the informal insurance they can obtain from their marriage. Chapter 1 proposes three forms of informal insurance: a marriage transfer from the groom which alleviates the bride's household credit constraints, an increase in the bride's household aggregate labour return once the groom joins the bride's household, and the socioeconomic network wife's family obtain from their daughter's marriage. While cultural marriage norms (such as bride price and matrilocality) and the social network marriage's gains are strong economic incentives for displaced households, they are not instrumental for women who stay at the origin since sending communities are credit-constrained after the event. These findings open the question of how policy could address the economic incentives for early marriage. Finally, Chapter 1 shows how unconditional cash transfers (UCT) can mitigate the marriage effects for IDPs. Understanding why forcibly displaced women and women of origin respond differently to economic incentives (i.e., cultural norms, social networks, future perceptions, and schooling) can contribute to effective policy design and evaluation.

Chapter 2 generates evidence of the long-term effect of hosting IDPs on displaced-hosting social participation and cohesion. IDP inflows have a long-term and sizable negative effect on displaced-hosting voter turnout and non-profit associations. Importantly, the larger the relative size of the forcibly displaced population in a municipality, the smaller the reduction in social participation. Chapter 2 documents that the decline in institutional trust and the decrease in inter-group trust between IDPs and natives are the underlying mechanisms behind the decrease in social participation. The proposed mechanisms imply that these findings have particular implications for settings where the source of the forced displacement has uneven impacts between natives and IDPs. Overall, by shedding light on the long-term effect of hosting IDP on decreasing social cohesion, Chapter 2 stresses the relevance of responding to the inter-group cohesion between native and forcibly displaced populations. Neglecting to respond to forcibly displaced population's integration may end up hurting social participation in host communities, with long-lasting consequences over the following decades.

Chapter 3 also examines the effect on hosting communities. In contrast to the second Chapter, it focuses on the short-term effects of IDP inflows. In particular, Chapter 3 examines the extent to which IDP inflows could pose challenges to eradicating infectious diseases. As we observed during the COVID pandemic, population movement could affect the spread of the virus. Chapter 3 produces suggestive evidence that IDP inflows increase the polio incidence in host communities, slowing polio eradication mechanically. Although potential endogeneity problems in Chapter 3 do not allow me to estimate the causal effect of IDP inflows, this Chapter is a first attempt to document that the arrival of non-vaccinated IDP children seems not to drive the increase in polio incidence. A vaccination program targeting forcibly displaced populations could be the policy behind these findings. Anecdotal and descriptive evidence suggests that the lack of extra efforts in reinforcing host communities' health capacity could potentially explain the effects. From a policy perspective, new efforts should be made to reach every child with vaccination and guarantee that locals and IDPs can access health services equally.

These three essays open avenues for future research. Inspired by the first Chapter, beyond the timing of marriage, I would like to understand better how displacement affects women and girls' behaviour before, during and after marriage. For instance, previous work in Egypt and Mexico shows how female genital mutilation (FGM) and bargaining power respond to income. However, it has not been explored how the income shock from forced displacement could affect the FGM of girls and the bargaining power of married women. Additionally, I would like to address questions on how marriage facilities network formation for displaced populations. Second, related to the second Chapter, I am interested in examining the economic, political and trust interactions in newly built villages between natives (or voluntary migrants) and the population displaced by reservoirs. This potential research project's biggest obstacle is identifying individual-level historical data. Finally, I would like to look at other relevant topics in health economics. Forced displacement affects depression, anxiety, and post-traumatic stress disorder, potentially impacting Inter Partner Violence. Generating more evidence on whether internal displacement affects Inter Partner Violence within IDP and native households in host communities could speak to the design and evaluation of policies targeting forcibly displaced households.

Bibliography

Abbas, Hassan (2008). “A Profile of Tehrik-I-Taliban Pakistan”. In: *CTC Sentinel. West Point, NY: Combating Terrorism Center. 1 (2): 1–4. Archived from the original on 1 January 2017. Retrieved 8 November 2008.*

Abel, M. (2019). “Long-Run Effects of Forced Resettlement: Evidence from Apartheid South Africa”. In: *The Journal of Economic History. Volume 79, Issue 4December 2019 , pp. 915-953.*

Acemoglu, D., A. Cheema, et al. (2020). “Trust in state and nonstate actors: Evidence from dispute resolution in Pakistan”. In: *Journal of Political Economy 128 (8):3090–3147.*

Acemoglu, D., T. Hassan, and J. Robinson (2011). “Social Structure and Development: A Legacy of the Holocaust in Russia”. In: *The Quarterly Journal of Economics, Volume 126, Issue 2, May 2011, Pages 895–946,*

Adda, Jérôme (2016). “Economic Activity and the Spread of Viral Diseases: Evidence from High Frequency Data”. In: *The Quarterly Journal of Economics, Volume 131, Issue 2, May 2016, Pages 891–941.*

Alix-Garcia, J. and D. Saah (2010). “The Effect of Refugee Inflows on Host Communities: Evidence from Tanzania”. In: *The World Bank Economic Review, Vol. 24, No. 1 (2010), pp. 148-170.*

Alix-Garcia, J., S. Walker, et al. (2018). “Do refugee camps help or hurt hosts? The case of Kakuma, Kenya”. In: *Journal of Development Economics. 130,66-83.*

Amat, F. and et.al (2020). “From Political Mobilization to Electoral Participation: Turnout in Barcelona in the 1930s”. In: *The Journal of Politics. Volume 82, Number 4. October 2020.*

Anderson, Siwan (2007). “The Economics of Dowry and Brideprice”. In: *Journal Economic Perspectives 21 (4): 151–74.*

Arbatli, C and G. Gokmen (2023). “Human Capital Transfers and Sub-national Development: Armenian and Greek Legacy in Post-expulsion Turkey”. In: *Journal of Economic Growth, 28, 1–43.*

Arroyo, W. and S. Eth (1995). “Assessment following violence-witnessing trauma”. In: *In E. Peled, P. G. Jaffe, J. L. Edleson (Eds.), Ending the cycle of violence: Community responses to children of battered women (pp. 27–42). Sage Publications, Inc.*

Asher, S. and et.al (2022). “The Long-run Development Impacts of Agricultural Productivity Gains: Evidence from Irrigation Canals in India”. In: *STEG Working Paper, June 2022.*

Ashraf, Nava et al. (2020). “Bride Price and Female Education”. In: *Journal of Political Economy. 2020; 128 (2), 591-641.*

Atkinson, M. and A. Fowler (2014). “Social Capital and Voter Turnout: Evidence from Saint’s Day Fiestas in Mexico”. In: *British Journal of Political Science, Vol. 44, No. 1 (JANUARY 2014), pp. 41-59.*

Baez, Javier E. (2011). "Civil wars beyond their borders: The human capital and health consequences of hosting refugees". In: *Journal of Development Economics 96 (2011) 391–408*.

Bahar, Dany, Ana-María Ibanez, and Sandra Rozo (2021). "Give me your tired and your poor: Impact of a large-scale amnesty program for undocumented refugees". In: *Journal of Development Economics Volume 151, June 2021, 102652*.

Banerjee, Abhijit et al. (2013). "Marry for what? Caste and mateselection in modern India". In: *American Economic Journal: Microeconomics, 5(2), 33–72*.

Bau, Natalie (2021). "Can Policy Change Culture? Government Pension Plans and Traditional Kinship Practices". In: *American Economic Review. Volume 111, No. 6, June 2021*.

Bauer, M. et al. (2016). "Can war foster cooperation?" In: *Journal of Economic Perspectives, 30(3): 249-274*.

Bazzi, Samuel, Sudarno Sumarto, and Asep Suryahadi (2015). "It's all in the timing: Cash transfers and consumption smoothing in a developing country". In: *Journal of Economic Behavior Organization, 2015, Volume 119, issue C, 267-288*.

Bechtel, M. and J. Hainmueller (2011). "How Lasting Is Voter Gratitude? An Analysis of the Short- and Long-Term Electoral Returns to Beneficial Policy". In: *American Journal of Political Science, Volume 55, Issue 4, October 2011, Pages 852-868*.

Becker, Gary (1973). "A Theory of Marriage: Part I". In: *J.P.E. 81 (4): 813–46*.

Becker, Sascha and Andreas Ferrara (2019). "Consequences of forced migration: A survey of recent findings". In: *Labour Economics Volume 59, August 2019, Pages 1-16*.

Becker, Sascha, Irena Grosfeld, et al. (2020). "Forced Migration and Human Capital: Evidence from Post-WWII Population Transfers". In: *American Economic Review. Volume 110, No. 5, May 2020*.

Bellows, J. and E. Miguel (2009). "War and Local Collective Action in Sierra Leone". In: *Journal of Public Economics 93 (11-12): 1144-1157*.

Bertoli, Simone, Elsa Gautrain, and Elie Murard (forthcoming). "Left behind, but not immobile: Living arrangements of Mexican transnational households". In: *Economic Development and Cultural Change, forthcoming*.

Bezhan, Faridullah (2014). "The Pashtunistan Issue and Politics in Afghanistan, 1947-1952". In: *Middle East Journal, Vol. 68, No. 2 (Spring 2014), pp. 197-209 (13 pages)*.

Bhalotraa, Sonia, Abhishek Chakravartyb, and Selim Gulesci (2020). "The price of gold: Dowry and death in India". In: *Journal of Development Economics Volume 143, March 2020, 102413*.

Bharadwaja, P. and R. Mirzab (2019). "Displacement and development: Long term impacts of population transfer in India". In: *Explorations in Economic History. Volume 73, July 2019, 101273*.

Blattman, C. and E. Miguel (2010). "Civil War". In: *Journal of Economic Literature 2010, 481, 3–57*.

Booth, A. et al. (2022). "The direct and intergenerational behavioural consequences of a sociopolitical upheaval". In: *Journal of Economic Behavior Organization, Volume 200, August, pp 931-958*.

Boserup, Ester (1970). "Woman's Role in Economic Development". In: *American Journal of Agricultural Economics, 1970, 283 pp*.

Bundervoet, Tom, Philip Verwimp, and Richard Akresh (2009). “Health and Civil War in Rural Burundi”. In: *Journal of Human Resources, March 2009, 44 (2) 536-563.*

Cagé, J. and V. Rueda (2016). “The Long-Term Effects of the Printing Press in sub-Saharan Africa”. In: *American Economic Journal: Applied Economics 2016, 8(3): 69–99.*

Calderon-Mejia, V. and A.M. Ibañez (2016). “Labour market effects of migration-related supply shocks: evidence from internal refugees in Colombia”. In: *Journal of Economic Geography, Volume 16, Issue 3, May 2016, Pages 695–713.*

Callaway, Brantly and PedroSant’Anna (2021). “Difference-in-Differences with multiple time periods”. In: *Journal of Econometrics. Volume 225, Issue 2, December 2021, Pages 200-230.*

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller (2008). “Bootstrap-based improvements for inference with clustered errors”. In: *The Review of Economics and Statistics, 90(3), 414–427.*

Carlana, Michela and Marco Tabellini (2020). “Happily Ever After: Immigration, Natives’ Marriage and Fertility”. In: *HBS Working Paper Series, 2020.*

Carlin.R, Love.G, and Zechmeister.E (2014). “Earthquake Damage, State Capacity, and Interpersonal Trust in Comparative Perspective”. In: *Comparative Politics 46(4): 419–37.*

Carpiano, R. (2008). “Actual or potential neighborhood resources and access to them: testing hypotheses of social capital for the health of female caregivers”. In: *Soc Sci Med. 2008 Aug;67(4):568-82. Epub 2008 Jun 10. PMID: 18547699.*

Castells-Quintana, David, Maria-del-Pilar Lopez-Uribe, and Thomas McDermott (2022). “Population displacement and urban conflict: Global evidence from more than 3300 flood events”. In: *Journal of Development Economics Volume 158, September 2022, 102922.*

CDC (2021). “Polio Vaccine: What you need to know”. In: *Centers for Disease Control and Prevention, 2021.*

Chaisemartin, Clément de and Xavier D’Haultfœuille (2020). “Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects”. In: *American Economic Review, Volume 110, NO. 9, September 2020, (pp. 2964-96).*

Chiappori, Pierre-André, Bernard Salanié, and Yoram Weiss (2017). “Partner choice and the marital college pre-mium: Analyzing marital patterns over several decades”. In: *AER. Volume 107, No. 8, August 2017.*

Chiovelli, Giorgio et al. (2021). “Forced Displacement, Human Capital and Occupational Choice”. In: *Working Paper.*

Chyn, Eric (2018). “Moved to Opportunity: The Long-Run Effects of Public Housing Demolition on Children”. In: *American Economic Review. Volume 108, no. 10, October 2018.*

Cole, S., A. Healy, and E. Werker (2012). “Do voters demand responsive governments? Evidence from Indian disaster relief”. In: *Journal of Development Economics, Volume 97, Issue 2, March 2012, Pages 167-181.*

Coleman, J. (1988). “Social Capital in the Creation of Human Capital”. In: *American Journal of Sociology Vol. 94, Supplement: Organizations and Institutions: Sociological and Economic Approaches to the Analysis of Social Structure (1988), pp. S95-S120 (26 pages).*

Corno, Lucia, Nicole Hildebrandt, and Alessandra Voena (2020). “Age of Marriage, Weather Shocks, and the Direction of Marriage Payments”. In: *Econometrica, May 2020, Volume 88, Issue 3.*

Depetris-Chauvin, E. and R. Santos (2018). “Unexpected guests: The impact of internal displacement inflows on rental prices in Colombian host cities”. In: *Journal of Development Economics Volume 134, September 2018, Pages 289-309.*

Deryugina, Tatyana, Laura Kawano, and Steven Levitt (2018). “The Economic Impact of Hurricane Katrina on Its Victims: Evidence from Individual Tax Returns”. In: *AEJ: Applied Economics, Volume 10No. 2 April 2018.*

Devkota, Bhimsen and Edwin R van Teijlingen (2010). “Understanding effects of armed conflict on health outcomes: the case of Nepal”. In: *Conflict and Health volume 4, Article number: 20 (2010).*

Din, Najam U (2010). “Internal Displacement in Pakistan: Contemporary Challenges”. In: *Human Rights Commission of Pakistan. Printed at Qasim Press Railway Road, Lahore.*

Duflo, E. and R. Pande (2007). “Dams”. In: *The Quarterly Journal of Economics, Volume 122, Issue 2, May 2007, Pages 601–646.*

Duranton, G. and M. Turner (2011). “The fundamental law of road congestion: Evidence from US cities”. In: *American Economic Review, Volume. 101, No. 6, October 2011, 2616-52.*

EiA (2018). “Grandes fracasos hidráulicos. Embalses carentes de utilidad”. In: *Ecologist in Action report.*

Engel, Stefanie and Ana-María Ibáñez (2007). “Displacement Due to Violence in Colombia: A Household-Level Analysis”. In: *Economic Development and Cultural Change, Vol. 55, No. 2 (January 2007), pp. 335-365 (31 pages).*

Fair, C. et al. (2017). “Natural Disasters and Political Engagement: Evidence from the 2010–11 Pakistani Floods”. In: *Quarterly Journal of Political Science 12:99–141.*

Fasani, Francesco, Tommaso Frattini, and Luigi Minale (2022). “(The Struggle for) Refugee integration into the labour market: evidence from Europe”. In: *Journal of Economic Geography, Volume 22, Issue 2, March 2022, Pages 351–393.*

Faye, Malang (2021). “A forced migration from Myanmar to Bangladesh and beyond: humanitarian response to Rohingya refugee crisis”. In: *Journal of International Humanitarian Action 6, 13 (2021).*

Fehr, E. (2009). “On the economics and biology of trust”. In: *Journal of the European Economic Association 7(2-3): 235-266.*

Flückiger, M., M. Ludwig, and A. Önder (2019). “Ebola and State Legitimacy”. In: *The Economic Journal, Volume 129, Issue 621, July 2019, Pages 2064–2089.*

Fowler, A. and A. Hall (2018). “Do Shark Attacks Influence Presidential Elections? Reassessing a Prominent Finding on Voter Competence”. In: *The Journal of Politics, Volume 80, Number 4.*

Frankenberg, Elizabeth and Lynn A. Karoly (1995). “The 1993 Indonesian Family Life Survey: Overview and Field Report”. In: *November, 1995. RAND, Santa Monica, CA.*

Frankenberg, Elizabeth and Duncan Thomas (2000). “The Indonesia Family Life Survey (IFLS): Study Design and Results from Waves 1 and 2”. In: *RAND, DRU-2238/1-NIA/NICHD.*

Gignoux, Jérémie and Marta Menéndez (2016). “Benefit in the wake of disaster: long-run effects of earthquakes on welfare in rural Indonesia”. In: *Journal of Development Economics. 118, 26–44.*

Goldin, Claudia (2006). “The Quiet Revolution That Transformed Women’s Employment, Education, and Family”. In: *A.E.R. 96 (2): 1–21.*

Gomez, B., T. Hansford, and G. Krause (2007). "The Republicans Should Pray for Rain: Weather, Turnout, and Voting in U.S. Presidential Elections". In: *The Journal of Politics, Vol. 69, No. 3 (Aug., 2007), pp. 649-663.*

Greenwood, Jeremy, Nezih Guner, and Guillaume Vandenbroucke (2017). "Family Economics Writ Large". In: *Journal of Economic Literature. Volume 55, No. 4, December 2017.*

Grossbard, Amyra (1978). "Towards a Marriage between Economics and Anthropology and a General Theory of Marriage". In: *A.E.R. 68 (2): 33–37.*

GTD (2021). "Global Terrorism Database". In: *Access at https://www.start.umd.edu/gtd/.*

Guiso, L. and et.al (2004). "The Role of Social Capital in Financial Development". In: *The American Economic Review, Vol. 94, No. 3 (Jun., 2004), pp. 526-556.*

Gunnsteinsson, Snaebjorn et al. (2022). "Protecting infants from natural disasters: The case of vitamin A supplementation and a tornado in Bangladesh". In: *Journal of Development Economics, Volume 158, September 2022, 102914.*

Hanaoka, Chie, Hitoshi Shigeoka, and Yasutora Watanabe (2018). "Do Risk Preferences Change? Evidence from the Great East Japan Earthquake". In: *American Economic Journal:Applied Economics, Volume 10, No. 2, April 2018 (pp. 298-330).*

Hetherington, M. (1998). "The Political Relevance of Political Trust". In: *The American Political Science Review, Vol. 92, No. 4, December 1998, pp. 791-808.*

Hussain, Shoaib F. et al. (2016). "Eradicating polio in Pakistan: an analysis of the challenges and solutions to this security and health issue". In: *Global Health. 2016 Oct 12;12(1):63.*

Ibáñez, Ana-María, Andrés Moya, et al. (2023). "Least Protected, Most Affected: Impacts of Migration Regularization Programs on Pandemic Resilience". In: *AEA Papers and Proceedings. 2023. Forthcoming.*

Ibáñez, Ana-María, Sandra Rozo, and María Urbina (2021). "Forced Migration and the Spread of Infectious Diseases". In: *Journal of Health Economics. Volume 79, September 2021, 102491.*

IDMC (2020). "Global Report on Internal Displacement". In: *Internal Displacement Monitoring Centre. Report.*

— (2017). "Global Report on International Displacement". In: *Geneva. Internal Displacement Monitoring Center Annual report.*

— (2015). "Solutions to displacement elusive for both new and protracted IDPs". In: *Geneva: Internal Displacement Monitoring Centre.*

Jones, Gavin W. (2004). "Not 'When to Marry' but 'Whether to Marry': The Changing Context of Marriage Decisions in East and Southeast Asia". In: *Untying the knot: Ideal and reality in Asian marriage, edited by Gavin W. Jones and K. Ramdas, 3-58. Singapore: NUS Press.*

Kijewski, S. and M. Freitag (2018). "Civil war and the formation of social trust in Kosovo: Posttraumatic growth or war-related distress?" In: *Journal of Conflict Resolution, 62(4), 717–742.*

Kirchberger, Martina (2017). "Natural disasters and labor markets". In: *Journal of Development Economics 125 (2017) 40–58.*

León-Ciliotta, G., D. Zejcirovic, and F. Fernandez (2023). "Policy-Making, Trust and the Demand for Public Services: Evidence from a Mass Sterilizations Campaign". In: *Working paper.*

Levy, M. (2018). "The Effect of Immigration from Mexico on Social Capital in the United States". In: *International Migration Review.*

Lu, Frances, Sameem Siddiqui, and Prashant Bharadwaj (2021). “Marriage Outcomes of Displaced Women”. In: *Journal of Development Economics. Volume 152, September 2021, 102684*.

Maffioli, E. (2021). “The political economy of health epidemics: Evidence from the Ebola outbreak”. In: *Journal of Development Economics, 2021, vol. 151, issue C*.

Malik, Adeel, Rinchan Ali Mirza, and Faiz Ur Rehman (2023). “Frontier Governmentality”. In: *Working Paper*.

Marcos, J. and M. Fernández (2019). “Memorias Ahogadas”. In: *Desplazados.org*.

Marsh, W. (2022). “Trauma and Turnout: The Political Consequences of Traumatic Events”. In: *American Political Science Review , December 2022, pp. 1 - 17*.

Martinez-Bravo, Monica and Andreas Stegmann (2022). “In Vaccines We Trust? The Effects of the CIA’s Vaccine Ruse on Immunization in Pakistan”. In: *Journal of the European Economic Association, Volume 20, Issue 1, February 2022, Pages 150–186*.

McGavock, Tamara (2021). “Here waits the bride? The effect of Ethiopia’s child marriage law”. In: *Journal of Development Economics. Volume 149, March 2021, 102580*.

Montalvo, Jose and Marta Reynal-Querol (2007). “Fighting against Malaria: Prevent wars while waiting for the "Miraculous" Vaccine”. In: *The Review of Economics and Statistics, February 2007, 89(1): 165–177*.

Moral, L. (2009). “Política hidráulica y desequilibrios territoriales en España. Historia y perspectivas. Agua, territorio y paisaje: de los instrumentos programados a la planificación aplicada”. In: *Sevilla: V Congreso Internacional de Ordenación del Territorio, 2009. 59-75, p. 61*.

Morales, J. (2018). “The impact of internal displacement on destination communities: evidence from the Colombian conflict”. In: *Journal of Development Economics Volume 131, March 2018, Pages 132-150*.

Mousa, S. (2020). “Building social cohesion between Christians and Muslims through soccer in post-ISIS Iraq”. In: *Science, 14 Aug 2020, Vol 369, Issue 6505,mpp. 866-87*.

Murard, E. and S. Sakalli (2021). “Mass refugee inflow and long-run Prosperity: Lessons from the Greek Population Resettlement”. In: *IZA Working Paper No. 11613*.

Murdock, George P. (1967). “Ethnographic Atlas”. In: *University Pittsburgh Press*.

Mushtaq, Asim et al. (2015). “Polio in Pakistan: Social constraints and travel implications”. In: *Travel Medicine and Infectious Disease (2015) 13, 360e366*.

Nakamura, Emi, Jósef Sigurdsson, and Jón Steinsson (2022). “The Gift of Moving: Intergenerational Consequences of a Mobility Shock”. In: *Review of Economic Studies, 2022*.

New-America (2021). “World of Drones”. In: *Access at https://www.newamerica.org/*.

Nisa, Eva (2016). “Marriage Practices: Indonesia”. In: *In S. Joseph (Ed.), Encyclopedia of Women Islamic Cultures Online*.

Nunn, N. and L. Wantchekon (2011). “The Slave Trade and the Origins of Mistrust in Africa”. In: *American Economic Review, 101 (7): 3221-52*.

OCHA (2022). “Internal Displacement”. In: *United Nations Office for the Coordination of Humanitarian Affairs, Access 21 May 2023. Access at https://www.unocha.org/internal-displacement*.

Oster, Emily (2012). “Routes of Infection: Exports and Hiv Incidence in Sub-Saharan Africa”. In: *Journal of the European Economic Association, Volume 10, Issue 5, 1 October 2012, Pages 1025–1058*.

Özler, Berk et al. (2021). “Children on the move: Progressive redistribution of humanitarian cash transfers among refugees”. In: *Journal of Development Economics Volume 153, November 2021, 102733*.

Pérez-Sindín, Xaquín, Tzu-Hsin Karen Chen, and Alexander Prishchepov (2021). “Are night-time lights a good proxy of economic activity in rural areas in middle and low-income countries? Examining the empirical evidence from Colombia”. In: *Remote Sensing Applications: Society and Environment. Volume 24, November 2021, 100647*.

Phadera, Lokendra (2021). “Unfortunate Moms and Unfortunate Children: Impact of the Nepali Civil War on Women’s Stature and Intergenerational Health”. In: *Journal of Health Economics. Volume 76, March 2021, 102410*.

Pinilla, V. and L. Sáez (2016). “Rural Depopulation in Spain: Genesis of a problem and innovative policies”. In: *Centre for Studies on Depopulation and Development of Rural Areas*.

Putnam, R. (1995). “Bowling Alone: America’s Declining Social Capital”. In: *Journal of Democracy 6, 1: 65-7*.

Rahim, Shabina, Zubair Ahmad, and Jamshid Abdul-Ghafar (2022). “The polio vaccination story of Pakistan”. In: *Vaccine. Volume 40, Issue 3, 24 January 2022, Pages 397-402*.

Robinson, Kathryn (2010). “Gender, Islam and democracy in Indonesia”. In: *Journal of Islamic Studies, Volume 21, Issue 3, September 2010, Pages 476–479*.

Rohner, D, M. Thoenig, and F. Zilibotti (2013). “Seeds of distrust: conflict in Uganda”. In: *Journal of Economic Growth, Vol. 18, No. 3 (September 2013), pp. 217-252*.

Romero, L. (2013). “La construcción de sociedades hidráulicas: el caso de España y del Oeste de EE.UU”. In: *Cuadernos de geografía, nº 93. Valencia: 2013. 53-77*.

Rozo, Sandra and Juan Vargas (2021). “Brothers or invaders? How crisis-driven migrants shape voting behavior”. In: *Journal of Development Economics 150 (2021) 102636*.

Sacerdote, Bruce (2012). “When the Saints Go Marching Out: Long-Term Outcomes for Student Evacuees from Hurricanes Katrina and Rita”. In: *American Economic Journal: Applied Economics. Volume 4, No. 1, January 2012*.

Saldarriaga, J. and Y. Hua (2019). “A gravity model analysis of forced displacement in Colombia”. In: *Cities Volume 95, December 2019, 102407*.

Sarsons, H. (2015). “Rainfall and Conflict: A Cautionary Tale”. In: *Journal of Development Economics [Internet]. 2015;115 (July) :62-72*.

Strauss, John, Firman Witoelar, and Bondan Sikoki (2016). “The Fifth Wave of the Indonesia Family Life Survey(IFLS5): Overview and Field Report”. In: *March 2016. WR-1143/1-NIA/NICHD*.

Sun, Liyang and Sarah Abraham (2021). “Estimating dynamic treatment effects in event studies with heterogeneous treatment effects”. In: *Journal of Econometrics. Volume 225, Issue 2, December 2021, Pages 175-199*.

Tabellini, G. (2008). “Institutions and Culture”. In: *Journal of the European Economic Association, Volume 6, Issue 2-3, 1 May 2008, Pages 255–294*.

Takasaki, Yoshito (2017). “Do Natural Disasters Decrease the Gender Gap in Schooling?” In: *World Development Volume 94, June 2017, Pages 75-89*.

Tertilt, Michèle (2005). “Polygyny, Fertility, and Savings”. In: *J.P.E. 113 (6): 1341–71.*

Tesfaw, Tadele Akalu (2022). “Internal Displacement in Ethiopia: A Scoping Review of its Causes, Trends and Consequences”. In: *Journal of Internal Displacement, Volume 12, Issue 1, January 2022, 2-31.*

Testa, P. (2021). “The Economic Legacy of Expulsion: Lessons from Post-War Czechoslovakia”. In: *The Economic Journal, Volume 131, Issue 637, July 2021, Pages 2233–2271.*

Tur-Prats, A. and F. Valencia-Caicedo (2020). “The Long Shadow of the Spanish Civil War”. In: *CEPR Working Paper.*

UN (1998). “Guiding Principles on Internal Displacement”. In: *United Nations Office for the Coordination of Humanitarian Affairs. United Nations.*

UNDESA (2015). “World Population Prospects: The 2015 Revision”. In: *United Nations Department of Economic and Social Affairs.*

UNFPA (2018). “Delivering Supplies When Crisis Strikes: Reproductive Health in Humanitarian Settings”. In: *United Nations Population Fund.*

UNHCR (2017). “Global trends. Forced Displacement in 2017”. In: *United Nations High Commis-sioner for Refugees.*

— (2022). “Global trends. Forced Displacement in 2021”. In: *United Nations High Commis-sioner for Refugees.*

UNICEF (2019). “National Emergency Action Plan 2018-2019”. In: *UNICEF.*

— (2023). “The State of the World’s Children 2023: For every child, vaccination”. In: *UNICEF Innocenti – Global Office of Research and Foresight, Florence, April 2023.*

— (2020). “UNICEF annual report 2019”. In: *United Nations Report.*

Vogl, Tom (2013). “Marriage Institutions and Sibling Competition: Evidence from South Asia”. In: *The Quarterly Journal of Economics, Volume 128, Issue 3, August 2013, Pages 1017-1072.*

Vogt, Manuel et al. (2015). “Integrating Data on Ethnicity, Geography, and Conflict: The Ethnic Power Relations Data Set Family”. In: *Journal of Conflict Resolution 59(7): 1327–42.*

Vroklage, Bernh (1952). “Bride Price or Dower”. In: *Anthropos 47 (1/2): 133–46.*

Wald, David J. et al. (2006). “ShakeMap Manual”. In: *United States Geological Survey.*

WCLD (2000). “Dams and Development a New Framework for decision-making”. In: *Earthscan Publications Ltd, London and Sterling.*

Weidmann, Nils B., Jan Ketil Rød, and Lars-Erik Cederman (2010). “Representing Ethnic Groups in Space: A New Dataset”. In: *Journal of Peace Research 47(4): 491–99.*

WHO (2022). “Polio vaccines: WHO position paper – June 2022”. In: *World Health Organization, 2022, 97, 277–300.*