Title

Essays on Healthcare Access and Demand for Healthcare in Low-Income Countries

ISBN

9798186163015

Author

Budillon, Fulvia

Publication Date

2026-07-29

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Essays on Healthcare Access and Demand for Healthcare in Low-Income Countries

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Economics

by

Fulvia Budillon

Committee in charge:

Professor Tom Vogl, Co-Chair
Professor Craig McIntosh, Co-Chair
Professor Prashant Bharadwaj
Professor Sara Lowes
Professor Paul Niehaus

2026

The Dissertation of Fulvia Budillon is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2026

DEDICATION

To mothers — may every mother have access to the care she needs to bring new life safely into the world.

To my mother.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank my advisors Craig McIntosh and Tom Vogl for their guidance and encouragement during this journey. They have been exceptional mentors, and I have learned so much from their intellectual honesty, curiosity, and kindness, values that I wish to bring into my own journey as an academic.

I am also extremely thankful to Prashant Bharadwaj, Sara Lowes, and Paul Niehaus for their invaluable feedback, stimulating discussions, and generosity with their time. I also thank my coauthors Sabareesh and Andre for teaching me so much and for learning together, for taking the lead when needed and for keeping up the enthusiasm when things don't seem to work.

I'm deeply grateful to Francesco, Paula, Andre, and Jordán for navigating together the early doctoral years during the pandemic, and for being a team in celebrating together in joy and helping each other in hardship.

I am also very grateful to the nurses and healthcare workers in the labor rooms I visited in Kenya and Mozambique, who generously shared their time and insights in answering my questions about their work.

Thanks to my friends, who have been my fresh air and my point of reference. You make my life brighter every day.

The warmest thanks to my family - mamma, papà, and Alberto - who have been my constant source of retreat, offering me love and unconditional support, even knowing how much it cost us to be apart. I owe you this and all my achievements, for the care you took to nourish my roots and to build my wings.

Chapter 2, in full, is currently being prepared for submission for publication of the material. Budillon, Fulvia; Ramachandran, Sabareesh. The dissertation author was one of the primary investigators and authors of this material.

VITA

2017	Bachelor of Science, University of Naples Federico II, Italy
2020	Master of Science, University of Naples Federico II, Italy
2026	Doctor of Philosophy, University of California San Diego

FIELDS OF STUDY

Major Field: Economics (Development Economics and Political Economy)

Studies in Development Economics
Professors Karthik Muralidharan, Paul Niehaus and Tom Vogl

Studies in Political Economy
Professors Renee Bowen and Sara Lowes

ABSTRACT OF THE DISSERTATION

Essays on Healthcare Access and Demand for Healthcare in Low-Income Countries

by

Fulvia Budillon

Doctor of Philosophy in Economics

University of California San Diego, 2026

Professor Tom Vogl, Co-Chair
Professor Craig McIntosh, Co-Chair

This dissertation consists of three chapters on healthcare access and demand for healthcare in low-income countries.

Chapter 1 studies the impact of expanding access to facility birth on early-life mortality and subsequent healthcare use in Malawi. Using an exogenous increase in travel costs to health facilities, I find that only high-risk pregnancies with access to high-quality care experience significant survival gains. The systematic mismatch between high-risk pregnancies and low-quality facilities at the last mile helps explain the limited mortality reductions from access expansion. Exposure to low-quality facilities also decreases future healthcare visits, suggesting

that expanding access to poor service quality can backfire by discouraging engagement with formal care.

In Chapter 2, we leverage the 5x increase in medical colleges in India between 1980 and 2020 to measure the impact of college openings in a district on access to care, take-up of care, and health outcomes. We find that one additional batch of graduating students is associated with a 4.4% increase in health facilities and a 6.7% increase in health workers in the district, concentrated almost entirely in the private sector in urban areas. The mechanism operates through students from the district enrolling in the local college and becoming doctors. This leads to an increase in healthcare access, but we do not detect improvements in morbidity or mortality. Our findings suggest that reforming public hiring to leverage location preferences of local graduates could expand the supply of doctors in the public sector too.

In Chapter 3, I examine how mothers in Sub-Saharan Africa update their delivery choices in response to day-of-birth deaths observed in their village. Using Demographic and Health Survey data from 38 countries and exploiting within-village variation in preceding births, I find that mothers shift toward facility delivery after home-birth deaths, while their response to facility-birth deaths depends on the observable risk of the pregnancy that died: they shift away from facilities when a low-risk pregnancy dies there, and toward higher-level facilities when a high-risk pregnancy dies there. These patterns are consistent with mothers attributing locally observed deaths to facility quality or pregnancy risk depending on the observable circumstances.

Chapter 1

Beyond Access: Facility Birth, Healthcare Use, and Early-Life Mortality in Malawi

1.1 Introduction

Sub-Saharan Africa continues to face the world's highest under-five mortality, with one in fifteen children dying before their fifth birthday (World Bank, 2023). Because half of these deaths occur within the first month of life, encouraging mothers to give birth in healthcare facilities rather than at home has been a central policy focus in recent decades as a means to reduce maternal and neonatal mortality. Yet, despite major gains in access, quality of care remains strikingly low, especially at the last mile where vulnerability is highest. When those with the greatest need for skilled care are served by the weakest facilities, expanding access alone may fail to deliver substantial health gains. Moreover, a poor experience with low-quality care can erode trust in the system, discouraging future use and perpetuating poor outcomes.

This paper addresses two questions. First, does access to facility birth reduce neonatal mortality, and how do effects vary jointly with facility quality and baseline risk at the last mile? Second, how does an exogenous encounter with formal healthcare affect mothers' future use of health services? Together, these questions examine both the immediate and dynamic consequences of expanding access when quality of care is uneven.

Despite extensive efforts to promote facility delivery, no study has demonstrated a clear causal effect of access on neonatal survival. Prior quasi-experimental analyses have failed to detect significant impacts (Okeke et al., 2015; Gabrysch et al., 2019; Shajarizadeh and Grépin, 2022; Jung and Kim, 2024). This paper provides the first causal evidence on the effects of access to facility delivery on early-life mortality, showing that previous null findings conceal substantial heterogeneity: expanding access saves lives only when high-risk pregnancies reach facilities that meet minimum standards of quality.

I combine georeferenced data from the Malawi Demographic and Health Survey (DHS), a census of health facilities from the Service Provision Assessment (SPA), and high-resolution rainfall data from the Climate Hazards Group InfraRed Precipitation with Station (CHIRPS). The SPA includes precise GPS coordinates and detailed information on facility infrastructure, staffing, and service availability, allowing me to measure not only how easily mothers can reach a facility but also what type of facility they can access. The empirical strategy exploits the interaction between rainfall in the two weeks before birth and distance to the nearest facility as an exogenous shock to travel costs. The idea is that rainfall damages roads and disproportionately reduces access for more remote households, who are more exposed to road deterioration, thereby limiting their ability to reach a facility for delivery. This variation is plausibly exogenous because localized rainfall close to the date of birth can affect health outcomes or subsequent healthcare use only through its impact on travel costs. To account for other channels through which weather may influence mortality, I control for longer-term rainfall patterns over the preceding year.

I find three main results. First, access to facility birth substantially lowers neonatal mortality among high-risk pregnancies when the nearest facility has a doctor, but has no benefit and can even increase mortality for low-risk births or where facilities lack qualified staff. A simple policy simulation shows that expanding access alone would bring little improvement in survival, whereas raising the quality of existing facilities would generate substantial gains. These findings indicate that the share of facility births, despite being widely used as a policy metric, is not a sufficient statistic for evaluating progress in neonatal health, and that meaningful progress requires ensuring that high-risk pregnancies take place in higher-quality facilities. Second, the quality of care shapes long-term engagement: mothers who give birth in low-quality facilities become less likely to have healthcare visits in the future, whereas those exposed to higher-quality facilities show no change in subsequent use. Third, delivery choices display strong persistence, as mothers tend to give birth in the same type of place, either at home or in a facility, as in their previous delivery.

The empirical strategy identifies the causal effect of access to facility birth. It improves upon approaches that rely solely on distance or rainfall by exploiting their interaction, since both could be individually endogenous. I control directly for the main effects of distance and rainfall, so their interaction isolates short-term variation in travel costs to facilities. Although the variation comes from highly localized rainfall close to delivery, a concern is that the shock may still correlate with other factors influencing health outcomes through mechanisms unrelated to access. This would threaten identification if these effects differ systematically between remote and more accessible areas. One possibility is that the shock captures earlier rainfall, which can affect health through illness or agricultural production differentially in more remote areas. To address this, I control for monthly rainfall in the twelve months preceding birth and its interaction with distance. Another concern is that the shock may be correlated with expected rainfall, as parents may time pregnancies accordingly, which would threaten identification if such planning differs by remoteness. I address this by controlling for expected rainfall around the birth date, computed as the five-year historical average for the same two-week window. Finally, to account for seasonal fluctuations and time-specific shocks that may jointly affect road access and health outcomes, I include grid-month and month-year fixed effects.

I begin by documenting that the shock in travel costs reduces the probability that rural mothers give birth in a health facility. The shock has no predictive power in urban areas, which account for about seven percent of the sample, so I limit the analysis to rural clusters. Moving from the 25th to the 75th percentile of travel costs decreases the probability of facility delivery by about 0.6 percentage points. This reduced access shifts deliveries away from skilled personnel such as doctors or nurses toward traditional birth attendants (TBAs), who are women with no formal medical training and who typically assist births within the community. The first stage is strong, robust across specifications, and invariant to the set of controls included.

I estimate the causal effect of facility access on neonatal mortality, allowing for heterogeneity by the quality of the nearest facility and by the mother's ex-ante risk of complications. Quality is proxied by whether the facility reports at least one doctor on staff in the SPA survey, and risk is defined as the predicted probability of neonatal death from pre-birth characteristics. The results reveal a striking pattern of complementarity between risk and quality. Moving from low to high access to facility delivery, defined as going from the 75th to the 25th percentile of travel costs induced by the shock, reduces neonatal mortality by 1.7 percentage points (about 56 percent of the mean) for high-risk births with access to a doctor, but increases mortality by 0.1 percentage points (7.8 percent) for low-risk births. These findings indicate that the benefits of facility delivery depend critically on both the baseline risk of the pregnancy and the quality of available care, suggesting that average effects conceal substantial heterogeneity across populations and settings.

The complementarity between risk and quality reveals a fundamental mismatch at the last mile: the pregnancies with higher returns from quality are often served by the weakest facilities. This helps explain why expanding access alone has yielded limited improvements in survival. To assess the magnitude of these unrealized gains, I simulate population-level changes in neonatal mortality under three counterfactual policies: (i) expanding access only, by moving all home births to existing facilities without doctors; (ii) improving quality only, by moving all births that currently occur in facilities without doctors to facilities with a doctor; and (iii) expanding both access and quality, by ensuring that all births occur in facilities with a doctor. The simulations show that expanding access alone would have negligible impact, whereas improving quality would reduce neonatal mortality by about 18 percent. Combining both margins yields only slightly larger gains, implying that improvements in quality account for nearly all of the potential reduction in mortality.

Next, I use the exogenous exposure to formal healthcare induced by the shock to study how a single interaction with the health system shapes future behavior and whether these effects differ by the quality of available care. I construct an index of facility resources combining all characteristics that may influence mothers' experiences, including cleanliness, availability of services, equipment, and supplies. I find that an increase in facility access leads to a decline in the likelihood that the mother visited a healthcare facility in the year before the survey by about one percentage point (roughly 1.3 percent of the mean) for mothers whose nearest facilities have low resources, whereas access to better-equipped facilities has no significant effect. These results highlight the potential drawback of contact with low-quality healthcare, which can discourage future engagement with the health system. I also find that the reduction in healthcare use extends beyond maternity and postnatal care, as the results remain unchanged when restricting the sample to children older than 14 months, for whom the recall period excludes postnatal visits.

Finally, I find persistence in delivery choices: an increase in access to facility birth raises the probability that a mother gives birth in a facility in subsequent deliveries by about one percentage point, regardless of facility quality. This also implies that mothers who are prevented from reaching a facility because of the shock are more likely to give birth at home for their next child. This persistence suggests that childbirth decisions are shaped more by established habits than by updated assessments of quality, and highlights the importance of interventions that promote initial facility use to prevent persistent patterns of home delivery.

These results contribute to several strands of literature. First, I contribute to the policy debate on the effectiveness of expanding access to facility birth as a strategy to reduce maternal and early-life mortality in low-income settings. This literature faces two main empirical challenges. Individuals self-select into facility birth based on characteristics correlated with health outcomes, limiting most studies to correlational evidence (Worku et al., 2013; Tura et al., 2013; Leslie et al., 2016). In addition, although early-life mortality is high in relative terms, it remains rare in absolute terms, which reduces statistical power. As a result, even studies that attempt causal identification often fail to reject economically meaningful effects, particularly when treating facility birth as a uniform intervention (Okeke et al., 2015; Gabrysch et al., 2019; Shajarizadeh and Grépin, 2022; Jung and Kim, 2024). An exception is Andrew and Vera-Hernández (2022), who study India's Janani Suraksha Yojana (JSY) program and find that promoting facility birth led to higher neonatal mortality when facilities became overcrowded. My work differs from theirs as I estimate the direct effect of facility birth, focusing on heterogeneity based on pre-treatment characteristics of last-mile users, whereas their analysis examines general equilibrium effects driven by post-treatment overcrowding. I contribute to this literature by explaining why previous causal studies find insignificant average impacts, showing that average effects mask substantial heterogeneity. I provide the first causal evidence on how the effect of facility delivery depends jointly on the baseline risk of the pregnancy and the quality of available care, showing that improvements in access yield large benefits only when adequately staffed facilities reach high-risk pregnancies.

Second, I contribute to the literature on repeated interactions with the health system by providing new evidence on two behavioral mechanisms: habit formation and learning. Habit formation, where past use lowers the cost of future use (Becker and Murphy, 1988; Allcott and Rogers, 2014; Fujiwara et al., 2016), is typically documented for behaviors with higher frequency such as handwashing (Hussam et al., 2022) or preventive care (Jones et al., 2024). I show that it also occurs for a low-frequency behavior like childbirth, which is notable because standard models would predict habits to form too slowly to be detectable when the behavior is infrequent. Prior work on learning in healthcare has shown that individuals update their beliefs about provider quality through repeated interactions or social learning within communities (Leonard, 2007; Adhvaryu, 2014). I contribute to this literature by showing that even a single, salient exposure to formal healthcare can shape subsequent health-seeking behavior, with effects expanding beyond the specific service used, suggesting broader learning about the formal health system.

Finally, this paper contributes to the broader literature on the determinants of low demand for formal healthcare in developing countries (Banerjee et al., 2010; Dupas et al., 2011; Dupas and Miguel, 2017; Lowes and Montero, 2021), among others. I provide new causal evidence that negative personal experiences with the health system can generate persistent disengagement from formal healthcare.

The paper proceeds as follows. Section 1.2 provides background on facility birth in Malawi, highlighting its relevance as a case study of last-mile healthcare delivery in sub-Saharan Africa. Section 1.3 outlines the empirical strategy. Section 1.4 presents the results. Section 1.5 concludes.

1.2 Challenges in expanding healthcare access

The World Health Organization recommends skilled attendance at birth to safeguard maternal and child health. Skilled attendance is defined as childbirth care provided by trained and accredited health professionals, including doctors, nurses, and midwives. To meet these guidelines, over the past decades, policymakers in low-income countries have invested heavily in expanding access to formal healthcare. These efforts have led to substantial improvements in health outcomes and increased utilization of services. Yet, a large gap in health indicators remains between low- and high-income countries. As shown in Figure **??**, there is a negative cross-country correlation between facility birth rates and neonatal mortality. But even countries with nearly universal facility birth in Sub-Saharan Africa still fall short of the health outcome levels seen in regions of the world with higher income. A substantial share of neonatal deaths in these countries may therefore still be preventable.

One reason is that in last-mile settings, the quality of care often remains too low to yield meaningful health gains. Many newly established facilities in rural areas suffer from poor quality, with staff and supply shortages especially severe in remote regions. Staffing is a particularly critical issue: there are just 0.2 physicians per 1,000 people in Sub-Saharan Africa, compared to 4.1 in the European Union and 3.6 in the United States (World Bank, 2025). Evidence confirms that staffing plays a central role in reducing mortality. Okeke (2023a) studies a policy experiment in Nigeria where doctors were randomly assigned to some health facilities, while others received mid-level providers instead. Mortality declined measurably in communities that received a doctor but not in those assigned mid-level providers, suggesting that clinical expertise is a key constraint.

In addition to limited quality, last-mile settings are also home to a higher concentration of high-risk pregnancies. Whether expanding access improves survival therefore depends not only on average quality, but on the match between the risk profile of users and the capability of the facilities they reach. The effect of access can differ sharply across risk groups: high-risk births may experience higher or lower returns to quality depending on how the two interact. Formally, this interaction can be represented by the cross partial

$$\frac{\partial^2 \text{Mortality}}{\partial \text{Risk}\, \partial \text{Quality}}.$$

If this cross partial is positive, higher risk attenuates the returns to quality. This may occur when high-risk pregnancies have such low counterfactual survival that even basic interventions yield large benefits. In this case, expanding access at the very last mile should be prioritized before quality upgrades. If it is negative, higher risk amplifies the returns to quality. This may happen if high-risk pregnancies require more specialized care, so only higher-quality facilities can effectively reduce mortality. In this case, targeting quality improvements where risk is highest would yield the greatest gains. The empirical analysis below examines which of these mechanisms is more relevant in this setting.

The limitations of access without quality raise a broader policy challenge. Expanding the use of formal care is necessary but not sufficient to achieve large health gains. The key policy question is no longer just how to increase the take-up of formal healthcare, but under what conditions and for whom healthcare use translates into better outcomes. Identifying these conditions is crucial for policy design: it allows governments and donors to move beyond coverage targets toward goals that explicitly incorporate quality of care and ensure that expanded access delivers real health improvements.

Facility birth in Malawi offers an ideal context to study these dynamics. Over the past twenty years, the country increased the share of births attended by skilled personnel from 56% in 2000 to 96% in 2023 (World Bank, 2025). This dramatic expansion followed several policies promoting facility birth, including the removal of user fees in mission health facilities in 2004 (Manthalu et al., 2016). Between 2004 and 2016, maternal mortality in Malawi declined from 851 to 451 per 100,000 live births, and neonatal mortality fell from 31 to 19 per 1,000 live births. While this represents substantial progress, these figures remain far above those observed in high-income regions with comparable facility coverage. In countries like Japan or Sweden, maternal mortality is around 4 per 100,000 and neonatal mortality as low as 1 per 1,000, suggesting that most of the remaining deaths are preventable with adequate care quality.

I draw on two nationally representative datasets: the Malawi Demographic and Health Survey (DHS), which provides detailed information on all births from a representative sample of women, including maternal and child health outcomes; and the Malawi Service Provision Assessment (SPA), which offers a census of health facilities with detailed data on infrastructure, staffing, and service availability. Linking these sources allows me to analyze both demand- and supply-side dimensions of childbirth care. Malawi is uniquely suited for this analysis because it is the only country in Sub-Saharan Africa that simultaneously meets two data requirements: the DHS records the exact day of birth, enabling a precise link between deliveries and short-run weather shocks in the empirical strategy, and the SPA provides full coverage of facilities, not a sample, offering comprehensive information on access and quality.

Despite this rapid expansion in coverage, access to highly skilled professionals remains scarce and unevenly distributed. Table A.1.2 shows that distance to the nearest facility is uncorrelated with neonatal mortality, while distance to the nearest facility with a doctor is negatively correlated with it: mortality is higher in clusters located further from a doctor. Figure 1.2 illustrates this disparity. Coverage appears widespread when all facilities are included, yet restricting to those with a doctor reveals a concentration in areas with relatively low neonatal mortality, whereas high-mortality clusters are typically far from any doctor. Neonatal deaths are thus concentrated where access to skilled professionals is most limited, reflecting both the causal effect of doctors in reducing mortality and their systematic allocation to better-off areas.

In this paper, I examine whether variation in two key dimensions - the quality of the facility and the baseline health risk of users - can explain why expanded access does not always lead to better health outcomes. First, I test whether the quality of available care shapes the extent to which access to facilities improves health, and which aspects of quality matter most. As shown in Table 1.2, quality varies markedly across facilities. SPA data indicate that rural hospitals and health centers, which are typically the most accessible for remote populations, have on average just 0.22 and 0.02 doctors per facility, respectively, compared to 33 in central hospitals and 2.87 in other hospitals. These last-mile facilities also lack basic infrastructure: only 95% of rural hospitals and 77% of health centers have running water onsite, while all central and other hospitals do.

Different elements of quality may affect outcomes through distinct channels. Clinical inputs such as staffing and functional equipment directly influence the medical effectiveness of care and hence the ability of facilities to improve survival. However, these inputs may have limited influence on future use, as rare outcomes like neonatal mortality are seldom observed in a single childbirth experience. By contrast, features related to infrastructure, cleanliness, and the availability of supplies may have a stronger effect on how users experience care and perceive the health system, precisely because these attributes are directly observed. I therefore examine how the effect of access to facility birth on neonatal mortality and on subsequent healthcare use varies along these two dimensions of quality, focusing on staffing when estimating effects on neonatal mortality, and on material resources when studying future healthcare use.

Second, I test whether access to facility birth is more beneficial for pregnancies with higher baseline risk and how this interacts with the quality of available care. Table 1.1 reports baby, mother, and village characteristics by place of delivery (healthcare facility or home[1]). DHS data show that mothers delivering in facilities differ systematically from those giving birth at home: they tend to be younger, more educated, and have fewer children; they also live closer to a facility, are less likely to be in rural areas, and are more likely to deliver in a facility with a doctor. Babies born in facilities have lower infant and neonatal mortality rates, are more likely to have received antenatal care, and show small differences in postnatal care. These patterns indicate strong selection into facility delivery: the women who already give birth in facilities

[1]The DHS does not collect place-of-delivery information for births occurring more than five years before the survey. These older births are systematically different: they tend to have lower birth order, higher mortality, and lower likelihood of visiting healthcare facilities, partly reflecting long-run trends of declining child mortality and increasing healthcare use. Detailed characteristics of the whole sample, including births for which place of birth is missing, are reported in Table A.1.3.

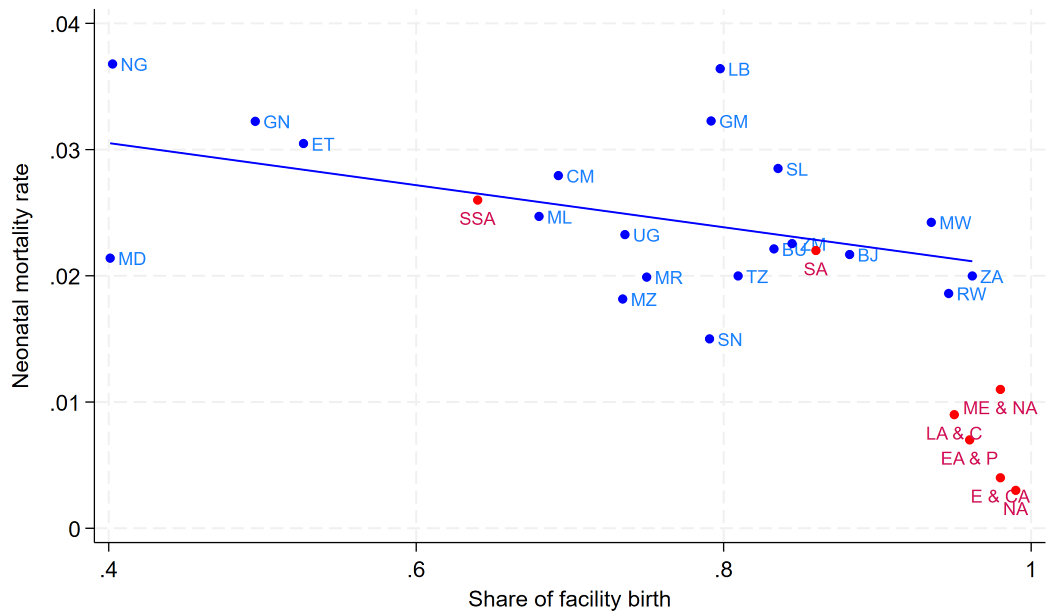


Figure 1.1. Neonatal mortality rate versus share of facility births, by country. Countries of Sub-Saharan Africa are shown in blue; world regions (East Asia and Pacific, Europe and Central Asia, Latin America and the Caribbean, the Middle East and North Africa, North America, South Asia, and Sub-Saharan Africa) are shown in red. Data for SSA countries are from the most recent available Demographic and Health Survey (DHS) for each country. Data for the world regions are from the World Bank Open Data (World Bank, 2025).

differ markedly from those who are reached by further access expansion. As access widens, the marginal users are therefore likely to be more disadvantaged and more vulnerable. In such settings, low-quality care may yield limited or even adverse returns if complications are not properly managed. This underscores the need to identify which dimensions of quality matter most, and for whom, so that policy can target improvements where they are likely to have the greatest impact.

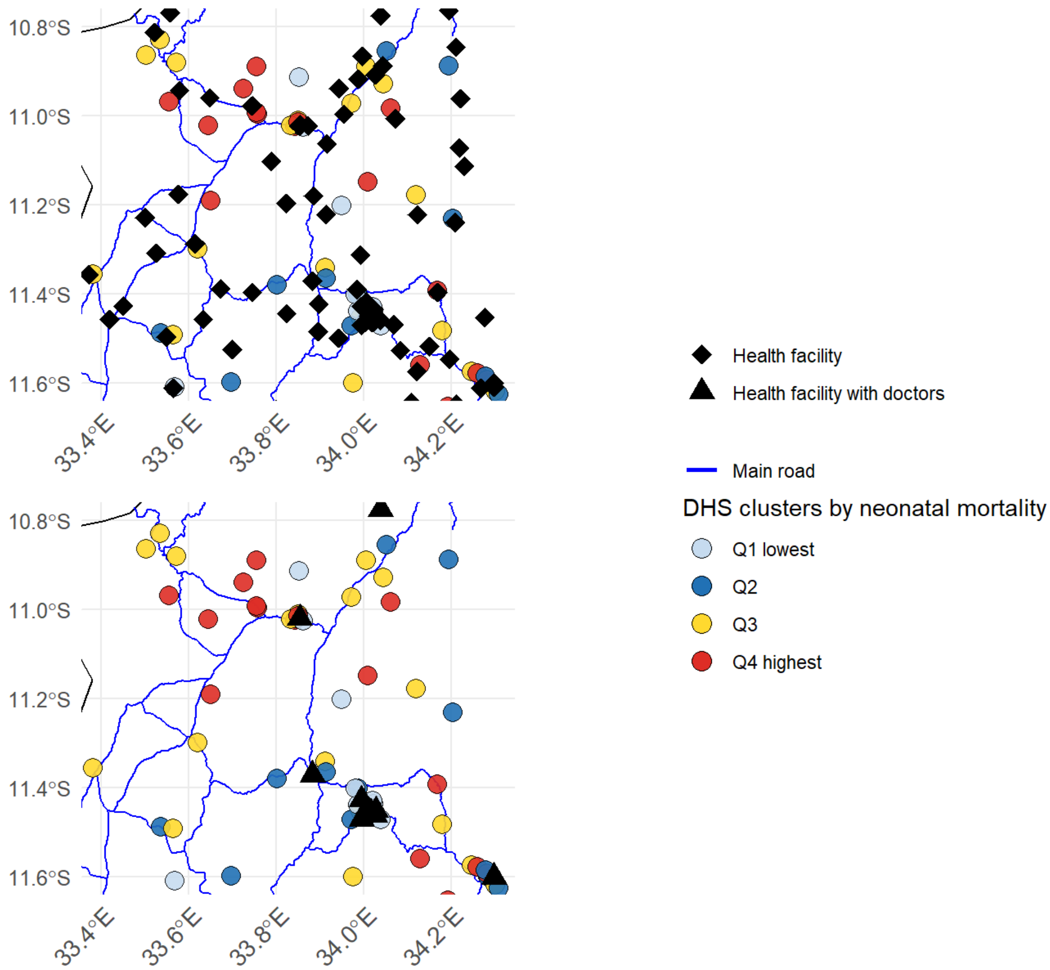


Figure 1.2. Distribution of healthcare facilities and neonatal mortality in Northern Malawi. The top panel shows all health facilities, while the bottom panel shows only those with a doctor. DHS clusters are colored by quartiles of neonatal mortality. Main roads are shown in blue.

Table 1.1. Baby, mother, and village characteristics by place of delivery. Column 1 reports mean values for births occurring at home, column 2 for births in facilities, and column 3 the difference between the two. Standard deviations are in parentheses.

	(1) Home	(2) Facility	(3) Difference (1) - (2)
Baby			
Birth order	3.88	3.14	0.74***
	(2.29)	(2.05)	(0.07)
Infant mortality (0/1)	0.05	0.04	0.02***
	(0.23)	(0.19)	(0.01)
Neonatal mortality (0/1)	0.04	0.02	0.02***
	(0.20)	(0.15)	(0.01)
Received antenatal care (0/1)	0.90	0.99	-0.09***
	(0.30)	(0.11)	(0.01)
Received postnatal care (0/1)	0.43	0.45	-0.02
	(0.49)	(0.50)	(0.02)
Mother			
Mother's age	27.01	25.91	1.10***
	(7.20)	(6.66)	(0.22)
Years of education	4.07	5.80	-1.73***
	(3.18)	(3.58)	(0.10)
Total children ever born	4.24	3.37	0.87***
	(2.31)	(2.08)	(0.07)
Visited health facility in last 12 months (0/1)	0.70	0.75	-0.04***
	(0.46)	(0.43)	(0.01)
Village			
Rural (0/1)	0.92	0.83	0.09***
	(0.26)	(0.37)	(0.01)
Distance to hospital (km)	9.25	6.71	2.54***
	(9.79)	(6.28)	(0.30)
Closest facility has a doctor (0/1)	0.15	0.21	-0.06***
	(0.35)	(0.40)	(0.01)
N	1,096	15,780	16,876

Table 1.2. Facility characteristics by type of health facility in Malawi.

	(1) Central Hospital	(2) District Hospital	(3) Rural Hospital	(4) Other Hospital	(5) Health Centre	(6) Maternity	(7) Dispensary	(8) Clinic	(9) Health Post	(10) All Facilities
General										
Rural (0/1)	0.00	0.12	0.93	0.27	0.94	1.00	0.91	0.36	0.96	0.68
	(0.00)	(0.34)	(0.26)	(0.45)	(0.25)	(0.00)	(0.29)	(0.48)	(0.19)	(0.47)
Has delivery services (0/1)	1.00	1.00	1.00	0.62	0.89	1.00	0.00	0.06	0.00	0.55
	(0.00)	(0.00)	(0.00)	(0.49)	(0.32)	(0.00)	(0.00)	(0.23)	(0.00)	(0.50)
Water onsite (0/1)	1.00	1.00	0.95	1.00	0.77	0.75	0.64	0.88	0.35	0.81
	(0.00)	(0.00)	(0.22)	(0.00)	(0.42)	(0.50)	(0.49)	(0.33)	(0.49)	(0.39)
Blood transfusion performed (0/1)	1.00	1.00	0.54	0.66	0.01	0.00	0.00	0.02	0.00	0.09
	(0.00)	(0.00)	(0.50)	(0.48)	(0.08)	(0.00)	(0.00)	(0.12)	(0.00)	(0.29)
Staffing										
Total staff	294.40	167.75	40.93	42.20	18.50	4.40	7.80	3.73	4.32	19.01
	(210.38)	(41.48)	(19.81)	(39.94)	(12.41)	(3.78)	(6.74)	(4.86)	(6.77)	(37.03)
Doctors	33.00	2.25	0.22	2.87	0.02	0.00	0.00	0.19	0.00	0.41
	(43.46)	(1.19)	(0.57)	(3.73)	(0.16)	(0.00)	(0.00)	(0.91)	(0.00)	(3.41)
Nurses	224.00	65.08	11.80	19.74	2.74	2.00	0.30	1.35	0.05	5.75
	(75.10)	(28.68)	(7.55)	(19.97)	(3.36)	(1.41)	(0.66)	(1.61)	(0.22)	(19.23)
Other health workers	111.00	100.42	28.90	21.38	16.17	3.50	8.83	2.81	6.00	14.46
	(41.09)	(21.39)	(14.06)	(22.07)	(9.98)	(3.32)	(6.52)	(3.69)	(7.37)	(19.56)
Patients										
Outpatient visits last month	.	6111.31	2279.89	1594.80	1593.20	597.00	1255.78	656.05	491.16	1345.46
	(.)	(3283.45)	(2041.87)	(1485.74)	(1249.05)	(784.25)	(1312.78)	(817.62)	(737.85)	(1449.92)
N	5	24	41	49	484	5	55	369	28	1,060

1.3 Empirical Strategy

1.3.1 Identification

This paper studies the causal effect of accessing facility birth on early-life mortality and subsequent healthcare use. A natural starting point would be an OLS regression comparing early-life mortality among children born in facilities to those born at home, and subsequent healthcare utilization among mothers who deliver in facilities versus those who deliver at home. However, such comparisons are likely to suffer from omitted variable bias due to nonrandom selection into facility delivery. Mothers choose whether to deliver in a facility based on characteristics that are also correlated with both their child's health outcomes and their own healthcare seeking behavior, confounding the estimated relationships.

A key concern is that this selection leads to nonzero covariance between facility birth and both health outcomes and healthcare use, biasing the estimated effect. This bias may be positive or negative, depending on the direction of selection. Positive selection arises if women who deliver in facilities are disproportionately those with higher socioeconomic status, and therefore better baseline health outcomes and greater healthcare use regardless of delivery location. For example, wealthier or more educated women, or those with better access to health-related information, may be more inclined to seek formal care and have children with lower risk of neonatal mortality. In this case, OLS would overstate the beneficial effect of facility birth. This pattern is well documented, especially in low income settings (Kruk et al., 2010; Thompson and Sofo, 2014; Andegiorgish et al., 2021).

Conversely, negative selection arises if women who deliver in facilities are more likely to have high risk pregnancies or worse underlying health. For instance, women with lower socioeconomic status may have limited routine engagement with the health system and children with poorer baseline health, but may nonetheless seek skilled care when complications are expected (Gowrisankaran and Town, 1999; Geweke et al., 2003; Doyle Jr, 2005; Adhvaryu and Nyshadham, 2010). In this case, OLS would understate the true effect of facility birth. Because both forms of selection are empirically plausible and may coexist within the same setting, the direction of the OLS bias is theoretically ambiguous.

To address concerns about selection bias, I use a shock to travel costs to healthcare facilities that generates exogenous variation in access to facility birth. The shock is the interaction between the mother's distance to the nearest delivery facility and rainfall around the day of birth. Specifically, I measure rainfall as the maximum daily rainfall in the two weeks before birth and distance as the length of the shortest road connection between the DHS cluster and the nearest health facility offering delivery services. The underlying logic is that rainfall worsens road conditions and raises travel costs more for women living farther from a facility. These women are more likely to be deterred from delivering in facility during periods of heavy rain, whereas women who live closer are less affected.

This approach improves on strategies that use distance or rainfall alone as shocks. Distance is a strong predictor of facility use but is clearly endogenous: households in remote areas differ systematically from those in better-connected locations in ways that are both observed and unobserved, including access to services, health behaviors, and socioeconomic status. Rainfall is plausibly exogenous but may influence outcomes through multiple channels unrelated to delivery location, for example, by affecting agricultural income, food security, or disease exposure. For either variable individually, the exclusion restriction is unlikely to hold.

The interaction of distance and rainfall addresses these concerns by capturing variation in temporary travel costs that affects facility access, while controlling for the direct effects of both distance and rain. In the empirical specification, I control for each main effect and use only the interaction as the instrument. The exclusion restriction is that, conditional on these controls, the interaction affects early-life mortality and subsequent healthcare use only through its impact on the probability of facility birth.

1.3.2 Addressing threats to identification

A potential concern is that the shock may capture differential effects of rainfall in more remote areas. Rainfall may influence agriculture and health differently depending on remoteness. For example, rain may be less valuable for agricultural production in areas farther from markets (Rogall, 2021), or more valuable in remote areas due to greater reliance on subsistence farming. Similarly, rainfall may have a stronger direct effect on health in remote areas due to poorer infrastructure, limited access to preventive care, or greater environmental exposure. To account for these channels, I control for monthly rainfall in the twelve months preceding birth, and their interaction with distance to the nearest facility. For instance, for a child born on June 10, 1995, I compute average daily rainfall over April 10-May 10, 1995, March 10-April 10, 1995, and so on until June 10-July 10, 1994, and interact each of these with distance.

Another threat to identification is that expected rainfall around the due date may influence childbirth timing differently across locations. More cautious parents - who also tend to have better health outcomes and higher healthcare use - may time pregnancies to avoid expected access constraints around delivery (Adhvaryu and Nyshadham, 2015). If these expectations matter more in remote areas, the exclusion restriction may be violated. To address this, I control for expected rainfall around the time of birth and its interaction with distance. Expected rainfall is defined as the average daily rainfall over a two-week window centered on the date of birth, computed over the five years preceding the birth. For example, for a child born on June 10, 1995, expected rainfall is the average rainfall from June 3 to June 17 across 1990–1994.

Seasonality also poses a threat, as both road conditions and health outcomes may vary systematically over the year. To account for location-specific seasonal patterns, I include grid-month fixed effects, where each grid is a 50 kilometer square containing about 10 DHS clusters. I also include month-year fixed effects to absorb time-specific national shocks, such as weather anomalies, that could jointly affect road quality and health outcomes.

1.3.3 Data

I combine several data sources. The primary source is the most recent round of the Malawi Demographic and Health Survey (DHS) collected during 2015–2016, a nationally representative household survey that collects detailed information on fertility, maternal and child health, and healthcare utilization. I use the birth recode file, where each observation corresponds to a child born to a woman interviewed in the survey. The dataset includes 65,005 births from 18,853 women. For the 16,876 births that occurred in the five years preceding the survey, the DHS collects detailed information on antenatal care, delivery, and postnatal care. Mortality and other outcomes are recorded for all births. Although earlier DHS rounds are available for Malawi, I restrict to this wave because it is the first to include day-of-birth information, which is essential for constructing daily rainfall measures around childbirth at precise temporal resolution. Earlier rounds report only the month of birth and cannot be used with this strategy.

To measure health facility characteristics, I use the Malawi Service Provision Assessment (SPA) collected during 2013-2014, which provides a census of all healthcare facilities in the country. The SPA includes GPS coordinates and detailed information on facility type, available services, staffing, equipment, and infrastructure, which allows for measuring both service availability and facility quality.

Rainfall data are from the Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS), which combines satellite imagery with ground station data to provide daily precipitation at a spatial resolution of 0.05 degrees (approximately 5.4 kilometers). I assign each DHS cluster to the nearest CHIRPS grid point based on geographic coordinates, and link individual births to daily rainfall patterns in the assigned grid around the date of birth.

To measure geographic access to care, I compute the distance from each DHS cluster to the nearest health facility offering delivery services. Distances are calculated using OpenStreetMap (OSM) data as the shortest travel route along the road network. Figure A.1.1 in the Appendix shows an example route from a village to its closest facility. Facilities that do not provide delivery care are excluded, ensuring that the distance measure reflects access to relevant childbirth services.

1.3.4 Estimating equation

I estimate the relationship between the shock in travel costs and facility birth using the following specification:

$$\begin{aligned} facility_{i,d,m,y,v,g} = &\alpha + \beta\ rain_{d,m,y,v,g} \times distance_{v,g} \\ &+ \gamma_1\ rain_{d,m,y,v,g} + \gamma_2\ distance_{v,g} \\ &+ \mathbf{X'}_{i,d,m,y,v,g}\theta + \delta_{m,g} + \tau_{m,y} + \varepsilon_{i,d,m,y,v,g} \end{aligned} \tag{1.1}$$

where $facility_{i,d,m,y,v,g}$ is a binary indicator equal to 1 if baby i born on $m/d/y$ in village v in grid g was born in a health facility, 0 if the baby was born at home; $rain_{d,m,y,v,g}$ is the maximum daily rain in village v within two weeks before the day of birth $m/d/y$; $distance_{v,g}$ is the road distance from village v to the closest health facility; $\mathbf{X'}_{i,d,m,y,v,g}$ is a vector of controls including average month-by-month rain in the year before birth, expected rain on the day of birth and their interaction with distance, as discussed in Section 1.3.2; $\delta_{m,g}$ is a seasonality (50 km grid-month) fixed effect; $\tau_{m,y}$ is a time (year-month) fixed effect; $\varepsilon_{i,d,m,y,v}$ is the error term.

I estimate the impact of the shock to travel costs on the outcomes of interest using the following specification:

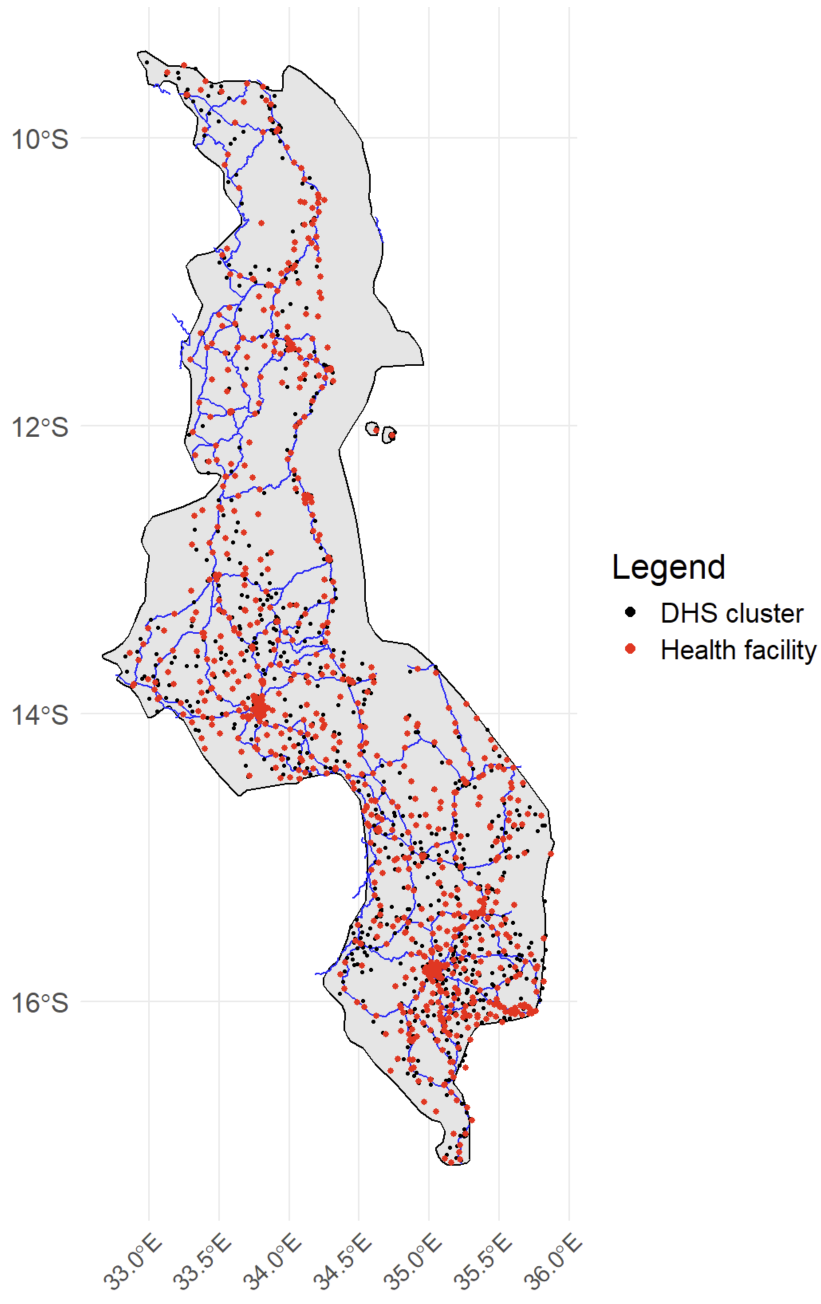


Figure 1.3. Geographic distribution of DHS clusters and health facilities. DHS clusters are shown as blue dots, and healthcare facilities from the Service Provision Assessment (SPA) are shown as red dots. Continuous lines indicate the paved road network.

$$
\begin{aligned}
y_{i,d,m,y,v,g} = & \alpha + \beta \ rain_{d,m,y,v,g} \times distance_{v,g} \\
& + \gamma_1 \ rain_{d,m,y,v,g} + \gamma_2 \ distance_{v,g} \\
& + \mathbf{X'}_{i,d,m,y,v,g} \theta + \delta_{m,g} + \tau_{m,y} + \varepsilon_{i,d,m,y,v,g}
\end{aligned} \tag{1.2}
$$

where $y_{i,d,m,y,v,g}$ represents one of the outcomes of interest: neonatal mortality, healthcare visits in the last 12 months, or facility birth for the next child.

My main specification relies on this reduced-form regression. The reason is that while place of birth is observed only for births that occurred in the five years preceding the survey, the shock and the outcomes are available for all births of each woman in the sample. The reduced form therefore increases the effective sample size by nearly a factor of four relative to a 2SLS approach, substantially improving statistical power and enabling a more precise analysis of heterogeneous effects. The identification of the reduced form still relies on the exclusion restriction that weather-induced travel costs affect mortality only through their impact on facility birth.

A central component of the analysis is to examine how the effect of the shock varies along two key dimensions: the quality of the available facility and the baseline health risk of the baby. In the reduced-form setting of equation (2), this is implemented by interacting the shock with categorical indicators for quality or baseline risk[2]. The resulting coefficients capture whether the same change in travel costs leads to different changes in outcomes for, for example, women whose nearest facility has a doctor versus those whose nearest facility does not, or for births classified as high- versus low-risk. This approach allows me to recover heterogeneous intention-to-treat effects of the shock to access on subpopulations.

As a complementary exercise, I also estimate a two-sample two-stage least squares (2S2SLS) specification in which the reduced-form instrument in equation (2) is replaced by the

[2] In all specifications, I also control for the main effects and all lower-order interactions between the components of the shock (rain and distance) and the heterogeneity categories. This ensures that the estimated coefficients on the triple interactions capture only the differential marginal effect of the access shock across groups, rather than level differences in exposure or baseline trends.

fitted value of facility birth from equation (1). In this framework, the rainfall–distance interaction serves as an instrument for facility birth, and β can be interpreted as the local average treatment effect for births whose delivery location is shifted by weather-induced changes in travel costs.

1.4 Results

1.4.1 First stage

Table 1.3 reports first-stage estimates of the effect of the interaction between rainfall and distance to the nearest delivery facility on the probability of facility birth. Results are presented for the full sample, as well as separately for rural and urban clusters. Standard errors are clustered at the grid-month level. The first-stage coefficients are negative and statistically significant, consistent with the interpretation that rainfall increases travel cost more for households located farther from a facility. However, this effect is entirely driven by the rural sample. In urban areas, the coefficient of the interaction term is not statistically different from zero, likely reflecting the fact that urban roads are paved and therefore less sensitive to rainfall. Based on these results, I restrict the main analysis to rural clusters, where the instrument is both conceptually valid and empirically relevant.

I also estimate a non-parametric version of the first stage, grouping both distance and rainfall into bins (Figure **??** in the Appendix). The pattern shows that the effect of the shock on facility birth is stronger at longer distances, and it is most pronounced under moderate rainfall levels, with little or no additional effect at extreme rainfall levels.

All specifications include controls for past rainfall and expected rainfall around the due date, as well as their interactions with distance. Seasonality and time fixed effects are also included. Table A.1.4 shows that the first stage relationship does not depend on the inclusion of controls and fixed effects. While the magnitude of the coefficient remains stable across specifications, precision improves with the inclusion of controls, suggesting that these adjustments absorb residual variation without driving the core result.

Table 1.3. First-stage results. The dependent variable is a binary indicator equal to 1 if the baby was born in a healthcare facility. Controls include past and expected rainfall, as well as their interactions with distance. Seasonality and time fixed effects are included. Standard errors are clustered at the grid-month level.

	Dep. var.: Facility birth (0/1)		
	(1)	(2)	(3)
	All	Rural	Urban
Rain (m) X Distance (5km)	-0.269***	-0.258***	1.700
	(0.0851)	(0.0897)	(1.149)
Rain (m)	0.00330	-0.0124	-0.757
	(0.251)	(0.270)	(0.917)
Distance (km)	-0.0428*	-0.0550***	-0.0278
	(0.0248)	(0.0121)	(0.0824)
N	13168	12106	984
N of clusters	758	748	213
Mean of Y	0.932	0.928	0.973
F (Kleibergen-Paap)	10.0	8.3	2.2
Effect p25 to p75	-0.00543	-0.00554	0.0129

Table 1.4. Reduced form estimates on assistance at birth. Controls include past and expected rainfall, as well as their interactions with distance. Seasonality and time fixed effects are included. Standard errors are clustered at the grid-month level.

	(1) Facility Birth (0/1)	(2) Skilled Attendant (0/1)	(3) Doctor (0/1)	(4) Nurse (0/1)	(5) TBA (0/1)	(6) Friends/ Relatives (0/1)
Rain (m) X Distance (5 km)	-0.258***	-0.322***	-0.122	-0.219	0.339***	0.0124
	(0.0897)	(0.0991)	(0.122)	(0.147)	(0.0968)	(0.0453)
N	12106	12307	12307	12307	12307	12307
N of clusters	748	754	754	754	754	754
Mean of Y	0.928	0.898	0.194	0.743	0.0330	0.0356
R^2	0.110	0.0975	0.0920	0.0968	0.104	0.0778
Effect p25 to p75	-0.00579	-0.00724	-0.00274	-0.00492	0.00762	0.000278

Table 1.4 provides further evidence on the interpretation of the shock. Column 1 replicates the first-stage estimate by reporting the effect of the instrument on the probability that the delivery occurred in a facility. The table then shows its effect on the likelihood that the delivery was attended by any skilled provider, and separately by a doctor, a nurse, a traditional birth attendant, or relatives and friends. Skilled attendance is defined as assistance by either a doctor, a nurse, or any accredited and trained personnel qualified to conduct deliveries. As expected, the decline in facility birth corresponds almost one-to-one with a decline in skilled attendance. Column 5 shows that mothers who do not deliver in a facility because of the shock are instead assisted by traditional birth attendants.

Falsification test

As a falsification test, I assess whether the first stage is truly capturing the increase in travel costs due to rainfall around the time of delivery, rather than any other confounder correlated with the mother's likelihood of giving birth in a facility. If this is the case, then the interaction between distance and rainfall occurring after birth, which cannot plausibly affect the delivery decision, should not predict facility use. Table 1.5 presents results from this test. Column 1 replicates the first stage, with the shock defined as the maximum rainfall in the two weeks preceding birth. Columns 2–4 report placebo specifications defined analogously but shifted forward in time, using the maximum rainfall in the two weeks before one, two, and three months

Table 1.5. Falsification test. The dependent variable is a binary indicator equal to 1 if the baby was born in a healthcare facility. The variables measuring rain one, two, or three months after birth are defined analogously to the shock rain variable: they capture the maximum daily rainfall in the two-week window occurring one, two, or three months after the birth date. Controls include past and expected rainfall, as well as their interactions with distance. Seasonality and time fixed effects are included. Standard errors are clustered at the grid-month level.

	Dep. var.: Facility birth (0/1)			
	(1)	(2)	(3)	(4)
Rain (m) × Distance (5km)	-0.258***	-0.253***	-0.252***	-0.261***
	(0.0897)	(0.0909)	(0.0879)	(0.0903)
Rain 1 month after birth (m) × Distance (5km)		-0.0698		
		(0.217)		
Rain 2 months after birth (m) × Distance (5km)			0.253	
			(0.156)	
Rain 3 months after birth (m) × Distance (5km)				-0.144
				(0.249)
N	12106	12106	12106	12106
N of clusters	748	748	748	748
Mean of Y	0.928	0.928	0.928	0.928
F (Kleibergen-Paap)	8.3	7.8	8.2	8.3
Effect p25 to p75	-0.00554	-0.00150	0.00543	-0.00308

after birth, respectively. In column 2, the coefficient is small and statistically insignificant. In column 3, it becomes large but positive and statistically insignificant. In column 4, the coefficient is smaller than the main first stage and negative, yet still too noisy to be distinguished from zero. Overall, these patterns indicate that the placebo shocks do not yield consistent or meaningful first-stage effects, reinforcing the interpretation that the main results are driven by weather shocks at the time of delivery.

1.4.2 Neonatal mortality

Table 1.6 reports estimates of the effect of access to facility birth on neonatal mortality, defined as death within 28 days of birth. I explore heterogeneity along two dimensions: staffing at the nearest facility and baseline risk of the delivery. For staffing, I construct an indicator for whether the nearest facility has a doctor. This choice is motivated by Okeke (2023a), who shows that assigning doctors, rather than mid-level providers, to facilities in randomly selected Nigerian communities reduces mortality. In my data, only 6.6 percent of facilities have at least one doctor, and just 23 percent of mothers live closest to a facility with a doctor.

The second source of heterogeneity is baseline mortality risk, proxied with a predicted mortality score based on pre-birth characteristics of the mother and child, such as maternal age, parity, and the baby's gender. Details on the construction of this index are provided in Section A.1.1 of the Appendix. I classify a delivery as high-risk if the predicted score is above the sample median, and low-risk otherwise.

Columns 1 and 2 present reduced-form estimates. Column 1 shows no significant average effect, in line with previous work attempting to identify the causal impact of facility birth on early-life mortality. Column 2 reveals that this masks substantial heterogeneity. Increased travel costs, which reduce access to facilities, substantially raise neonatal mortality for high-risk deliveries when the nearest facility is staffed with a doctor. This suggests that facility birth significantly lowers mortality for high-risk pregnancies when more highly trained staff are available. By contrast, for low-risk deliveries, higher travel costs lower neonatal mortality, consistent with the idea that facility care of low quality may be unnecessary or even harmful when complications are unlikely. Table A.1.5 complements these results by reporting reduced-form regressions in which the interaction terms are entered individually.

To gauge magnitudes, moving from the 25th to the 75th percentile of the shock increases mortality by 1.73 percentage points for high-risk mothers with access to a doctor, from 3.11 to 4.84 percent, a 56 percent increase. For low-risk deliveries, increased travel costs reduce mortality by 0.14 percentage points, or 7.8 percent relative to a baseline of 1.78 percent. These patterns suggest that avoiding low-quality care in low-risk cases may reduce harm.

The identification strategy does not allow the effect of facility care to be separated from the effect of travel. The observed decline in mortality when travel costs increase could reflect poor quality of care, risks from the journey, or both. Nevertheless, the results indicate that benefits for low-risk deliveries are too limited to outweigh even potential travel-related harms.

Columns 3–8 report estimates for the restricted sample of children born in the five years preceding the survey. For these births, place of delivery is observed, allowing estimation of a first stage (columns 3–4) of the effect of the shock on the probability of facility birth, and IV estimates of the effect of facility birth on neonatal mortality. Columns 7–8 report reduced-form estimates, analogous to columns 1–2, for this subsample.

The 2SLS coefficient in column 5 is positive and large in magnitude but statistically insignificant due to wide standard errors. Column 6 reports IV estimates allowing for heterogeneity along the two dimensions. Point estimates are again large but imprecise. Low statistical power reflects both the rarity of neonatal mortality, which makes detecting significant effects difficult even when baseline rates are high by international standards, and the smaller sample size.

1.4.3 Future healthcare use

I next turn to the effects of access to facility birth on subsequent healthcare use. I analyze two outcomes: whether the mother visited a health facility in the 12 months preceding the survey, and whether her next child was delivered in a facility. To explore heterogeneity, I construct an index of facility resources that captures observable characteristics likely to shape women's experiences and beliefs about the healthcare system. Unlike neonatal mortality, where clinical expertise plays a central role, what matters here is likely what mothers see and interact with during their visit. Most are unlikely to know the qualifications of attending staff, but may form lasting impressions based on the availability of supplies, equipment, and services. The resources index aggregates information on delivery and newborn supplies, equipment, medication, and services available at the nearest facility. Further details on how the index is constructed are provided in section A.1.1 of the Appendix.

Healthcare visits

Table 1.7 reports second-stage estimates of the effect of facility birth on subsequent healthcare visits. The outcome variable is a binary indicator equal to 1 if the mother reported visiting a healthcare facility in the 12 months preceding the survey. Because the outcome

Table 1.6. Second-stage results for neonatal mortality. The dependent variable is a binary indicator equal to 1 if the baby died within 28 days of birth. High risk is a dummy equal to 1 if the standardized log-odds from a logistic regression of neonatal mortality on pre-birth covariates is above the median. Doctor equals 1 if the closest facility has at least one doctor on staff. Controls include past and expected rainfall, as well as their interactions with distance. Seasonality and time fixed effects are included. Standard errors are clustered at the grid-month level.

	All births		Last 5 years births					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Reduced form *Neonatal Mortality (0/1)*		**First stage** *Facility birth (0/1)*		**IV** *Neonatal Mortality (0/1)*		**Reduced form** *Neonatal Mortality (0/1)*	
Rain (m) × Distance (5km)	-0.0147	-0.0643*	-0.280***	-0.236**			-0.0219	-0.0221
	(0.0288)	(0.0363)	(0.0877)	(0.0992)			(0.0382)	(0.0357)
Rain (m) × Distance (5km) × High risk = 1		0.109*		-0.673*				-0.0397
		(0.0589)		(0.362)				(0.156)
Rain (m) × Distance (5km) × Doctor = 1		0.126		0.826				-0.0903
		(0.158)		(0.740)				(0.259)
Rain (m) × Distance (5km) × High risk = 1 × Doctor = 1		0.629*		-1.339				0.579
		(0.348)		(2.011)				(1.186)
Facility birth (0/1)					0.0784	0.0831		
					(0.143)	(0.171)		
Facility birth (0/1) × High risk = 1						-0.00434		
						(0.188)		
Facility birth (0/1) × Doctor = 1						-0.316		
						(0.470)		
Facility birth (0/1) × High risk = 1 × Doctor = 1						-0.0688		
						(1.002)		
N	46567	46567	12106	12106	12106	12106	12106	12106
N of clusters	811	811	748	748	748	748	748	748
Mean of Y	0.026	0.026	0.928	0.928	0.025	0.025	0.024	0.024
Effect p25 to p75	0	-.001 *	-.006 ***	-.005 **			0	0
Effect p25 to p75 × High risk = 1		.001		-.019 ***				-.001
Effect p25 to p75 × Doctor = 1		.001		.013				-.002
Effect p25 to p75 × High risk = 1 × Doctor = 1		.017 **		-.03				.009
F (Kleibergen-Paap)					10.166	1.515		

is measured at the mother level while the shock is defined at the birth level, I weight each observation by the inverse of the number of births the mother had. I exclude births that occurred in the past 12 months to avoid counting facility births as part of subsequent healthcare visits.

Columns 1 and 2 present reduced-form estimates. Column 1 shows no significant average effect. Column 2 uncovers meaningful heterogeneity. An increase in travel costs reduces facility birth, which in turn raises the probability of healthcare visits among women whose nearest facility has limited resources. This is consistent with the idea that negative interactions with low-quality care may discourage future use. The coefficient on the interaction between the shock and high resource is negative and statistically significant. However, a test of the joint effect of the shock and its interaction with high resource is not significant. This indicates that the significance of the interaction term only reflects that the effect for women with access to high-resource facilities is not significantly different from zero, rather than being negative and statistically different from zero.

In terms of magnitude, moving from the 75th to the 25th percentile of travel costs, meaning going from low to high access to facility birth, decreases the probability of a healthcare visit in the last 12 months by about 1 percentage point for mothers whose nearest facility has limited resources. This corresponds to roughly 1.3 percent of the sample average (0.67).

Columns 5 and 6 report 2SLS estimates. The signs of the coefficients are consistent with the reduced-form results, but the estimates are imprecise, underscoring that the 2SLS specification lacks statistical power due to the smaller sample size.

A natural question is whether these visits relate mainly to maternity care or to other services. To address this, I further restrict the sample to babies at least 14 months old, which removes from the recall period postnatal care visits that typically occur within two months of birth. Results in Table A.1.6 remain significant, indicating that the effects extend to healthcare services other than delivery or postnatal care of the same baby.

Table 1.7. Second-stage results for future healthcare visits. The dependent variable is a binary indicator equal to 1 if the mother reported visiting a healthcare facility in the past 12 months. The outcome is measured at the mother level and observations are weighted by the inverse of the number of births per mother. Resources is a composite index of facility characteristics, grouped into terciles. Controls include past and expected rainfall, as well as their interactions with distance. Seasonality and time fixed effects are included. Standard errors are clustered at the mother level.

	All births		Last 5 years births					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Reduced form *Visited health facility last 12 months (0/1)*		**First stage** *Facility birth (0/1)*		**IV** *Visited health facility last 12 months (0/1)*		**Reduced form** *Visited health facility last 12 months (0/1)*	
Rain (m) × Distance (5km)	0.181	0.393***	-0.372***	-0.309**			0.137	0.439**
	(0.121)	(0.134)	(0.111)	(0.135)			(0.210)	(0.188)
Rain (m) × Distance (5km) × Resources = Intermediate		-0.706**		-0.0995				-1.030**
		(0.284)		(0.296)				(0.443)
Rain (m) × Distance (5km) × Resources = High		-0.542		0.0899				-0.432
		(0.355)		(0.426)				(0.808)
Facility birth (0/1)					-0.370	-0.468		
					(0.542)	(0.607)		
Facility birth (0/1) × Resources = Intermediate						-0.0185		
						(0.0439)		
Facility birth (0/1) × Resources = High						0.00699		
						(0.0731)		
N	43850	43850	9420	9420	9420	9420	9420	9420
N of clusters	12154	12154	7983	7983	7983	7983	7983	7983
Mean of Y	0.668	0.668	0.925	0.925	0.730	0.730	0.730	0.730
Effect p25 to p75	.004	.009 ***	-.008 ***	-.007 **			.003	.01 **
Effect p25 to p75 × Resources = Intermediate		-.007		-.009				-.013
Effect p25 to p75 × Resources = High		-.003		-.005				0
F (Kleibergen-Paap)					11.232	3.041		

Table 1.8. Second-stage results for subsequent facility birth. The dependent variable is a binary indicator equal to 1 if the next baby from the same mother was born in a healthcare facility. Resources is a composite index of facility characteristics, grouped into terciles. Controls include past and expected rainfall, the instrument measured at the time of the next birth, and their interactions with distance. Seasonality and time fixed effects are included. Standard errors are clustered at the mother level.

	All births		Last 5 years births					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Reduced form *Next baby* *Facility birth (0/1)*		**First stage** *Facility birth (0/1)*		**IV** *Next baby* *Facility birth (0/1)*		**Reduced form** *Next baby* *Facility birth (0/1)*	
Rain (m) × Distance (5km)	-0.393***	-0.432***	-0.397*	-0.323			-0.548***	-0.564***
	(0.128)	(0.120)	(0.241)	(0.248)			(0.203)	(0.214)
Rain (m) × Distance (5km) × Resources = Intermediate		0.196		0.0826				0.527
		(0.449)		(0.952)				(0.631)
Rain (m) × Distance (5km) × Resources = High		-0.0577		-0.657				-0.456
		(0.412)		(1.375)				(0.973)
Facility birth (0/1)					1.380	1.525*		
					(0.863)	(0.918)		
Facility birth (0/1) × Resources = Intermediate						-0.597		
						(1.321)		
Facility birth (0/1) × Resources = High						-0.541		
						(1.606)		
N	8749	8749	2368	2368	2368	2368	2368	2368
N of clusters	7075	7075	2274	2274	2274	2274	2274	2274
Mean of Y	0.920	0.920	0.899	0.899	0.929	0.929	0.929	0.929
Effect p25 to p75	-.008 ***	-.009 ***	-.01 *	-.008			-.013 ***	-.014 ***
Effect p25 to p75 × Resources = Intermediate		-.005		-.006				-.001
Effect p25 to p75 × Resources = High		-.01		-.024				-.025
F (Kleibergen-Paap)					2.722	0.279		

Subsequent facility birth

Table 1.8 reports second-stage estimates of the effect of facility birth on the probability that the subsequent child is born in a healthcare facility. The outcome variable is a binary indicator equal to 1 if the next baby from the same mother was delivered in a facility.

Column 1 shows a significant negative relationship between travel costs and subsequent facility birth. This result points to strong persistence in childbirth decisions: mothers who deliver in a facility are more likely to return to a facility for the next birth, while those who deliver at home tend to continue delivering at home. Moving from the 25th to the 75th percentile of the instrument decreases the probability of facility birth for the subsequent child by about 1 percentage point from a sample average of 92 percent.

Column 2 examines whether this persistence varies by facility resources. The interactions with the resources index, however, yield no statistically significant differences. This suggests that the persistence is not primarily driven by perceived quality but may instead reflect broader cultural and behavioral dynamics, consistent with habit formation in childbirth care decisions.

Column 5 reports a positive and nearly statistically significant 2SLS estimate. Column 6 includes interactions with quality terciles, but the estimates remain imprecise and non-significant. As in the other outcomes, the limited sample size constrains statistical power, making it difficult to estimate IV coefficients with precision.

The analysis of subsequent facility birth is restricted to mothers who have at least one additional child observed in the data. A potential concern is that this selected sample might not be random if the likelihood of having another birth is correlated with the instrument. In that case, the second-stage estimates could be biased because the instrument would affect the probability of entering the sample, violating the exclusion restriction. To assess this, I test whether the instrument predicts future fertility in Table A.1.7. The coefficient is not statistically significant in the full sample. I also estimate the effect restricting the analysis to (i) births that occurred at least three years before the survey and (ii) mothers aged 40 or older, two ways to condition on realized fertility, and find no significant relationship in either case. This reassures that the instrument does not predict whether a mother has another child, and that the estimated effects are not driven by endogenous sample selection.

1.4.4 Robustness checks

Alternative baseline mortality index

To assess whether the results are driven by socioeconomic factors rather than by clinical risk, I construct an alternative baseline mortality index that predicts neonatal mortality using only clinical characteristics of the birth. Specifically, the index is estimated using the year and month of birth, the baby's gender, the mother's age, birth order, birth spacing, whether the mother has ever had a terminated pregnancy, and the household's main source of drinking water (Bove et al., 1995, 2002). This specification aims to capture biological and reproductive risk while limiting the influence of socioeconomic and geographic variables. Although some of these factors, such as terminated pregnancies or water source, may still correlate with socioeconomic status, the exercise provides a useful robustness check that abstracts from most non-clinical sources of variation. The results remain robust to this alternative index, as shown in Table A.1.8, where the estimated effects on neonatal mortality are consistent in both magnitude and significance.

As an additional robustness check, I re-estimate the baseline mortality index using a linear probability model instead of the logistic specification used in the main analysis. The set of predictors remains identical to that of the main index. This exercise tests whether the results are sensitive to the functional form used to generate the mortality predictions. As shown in Table A.1.9, the estimated effects remain stable, indicating that the main findings are not driven by the choice of the underlying model for baseline mortality risk.

Alternative resources index

As a robustness check, I recompute the facility resources measure using an additive index defined in Section A.1.1 of the Appendix. Conceptually, MCA assigns higher weights to response categories that explain more of the variation across facilities, measured using a chi-square distance, whereas the additive index assigns equal weight to each criterion. I re-estimate all specifications substituting the additive resources index for the MCA resources index. The results, reported in Tables A.1.10 and A.1.11, closely match the baseline in sign and magnitude, and the main conclusions are unchanged.

Potential misreporting of the day of birth

This strategy relies on the reported day of birth, which may be subject to misreporting. Figure **??** shows the distribution of reported birth days. In the absence of misreporting, we would expect a roughly uniform distribution across days of the month (with days such as 29, 30, and 31 being slightly less frequent). However, the data display pronounced spikes on certain ”heaping” days such as the 1st, 10th, 12th, 15th, and 20th, suggesting systematic rather than random reporting errors. Table A.1.12 provides supporting evidence, showing that births reported on heaping days differ in observable characteristics from those reported on other days, with heaping being more prevalent among lower-SES mothers. As a robustness check, I exclude births reported on heaping days and re-estimate the main specifications. Results remain qualitatively unchanged, indicating that potential misreporting of the day of birth does not drive the main findings.

1.4.5 Alternative heterogeneity measures

For completeness, I also examine whether the results hold when heterogeneity is defined along the alternative quality dimension for each outcome. For neonatal mortality, I replace staffing, which directly captures the facility’s ability to respond to life-threatening complications, with the resources index described in Section A.1.1, which aggregates delivery and newborn supplies, equipment, medication, and services offered (Table A.1.13). While this broader index undoubtedly reflects many aspects that can improve the hospital experience, these dimensions may not translate into lower mortality if they are not directly related to managing obstetric emergencies. I therefore do not expect to find heterogeneity by this measure, and as expected, the results show no statistically significant effects.

For future healthcare use, I reverse the exercise, replacing the resources index with the doctor indicator (Table A.1.14 and Table A.1.15). Here I also do not expect to find heterogeneity, as mothers are often unaware of the formal training of their birth attendant and unlikely to distinguish between a doctor and a nurse. Instead, visible aspects of care, such as the availability of equipment, supplies, and services, are more likely to shape their delivery experience and influence subsequent healthcare behavior. As expected, no significant effects emerge.

The main takeaway is that quality investments should target the dimension most relevant for the outcome of interest: staffing capacity is key for reducing mortality, while facility infrastructure, supplies, and service availability are more influential for fostering sustained healthcare utilization. It is noteworthy that mothers appear to base their future use decisions on aspects of care that are largely irrelevant for survival in obstetric emergencies, such as equipment or the range of services offered, because these features shape the overall delivery experience. While such dimensions may not improve survival directly, investing in them remains important for retention, as they can strengthen trust in the health system and encourage continued engagement with formal care.

1.4.6 Back of the envelope calculations

This exercise translates the heterogeneous treatment effects into simple, policy-relevant magnitudes and compares scenarios that target different policy margins. The results show that the causal effect of facility delivery is large for high-risk births when they have access to adequately equipped facilities, while aggregate gains remain modest because many high-risk births are not matched to sufficient quality of care. I quantify how much neonatal mortality could fall under three counterfactual policies: expanding access alone, improving quality alone, and doing both. Calculations combine population shares and subgroup mortality in Table 1.9 with coefficients from Table A.1.5. I treat those coefficients as causal for this back-of-the-envelope exercise, assume no behavioral or general equilibrium responses, and assume the doctor effect does not vary across locations. Heterogeneity estimates may also capture selection or placement; I abstract from those concerns here to focus on relative magnitudes across policy margins.

Policy 1: Expanding access (quantity). Move all home births to facilities without a doctor. Among home births, 4.07% are low risk and 2.42% are high risk. The estimated effect for low-risk home $\rightarrow$ facility without doctor is $+0.001$, while for high-risk home $\rightarrow$ facility without doctor it is -0.001. The net change in the national mortality rate is:

$$\Delta_1 = (0.0242 \times 0.001) + (0.0407 \times (-0.001)) = 0.0000242 - 0.0000407 = -0.0000165$$

$$\text{New average} = 0.02424 - (-0.0000165) = 0.02426$$

This corresponds to a tiny increase of about 0.02 deaths per 1,000 births (+0.07%). The effect is essentially zero because the home-birth share is small and the estimated effects are very small and offsetting.

Policy 2: Improving quality. Give a doctor to all facility births currently in facilities without one. The relevant share is 27.23% of births. For low-risk births in these facilities (47.08%), the estimated effect of adding a doctor is zero. For high-risk births in these facilities, the estimated effect is -0.016. The change is:

$$\Delta_2 = 0.2723 \times (-0.016) = -0.00436$$

$$\text{New average} = 0.02424 - 0.00436 = 0.01988$$

The reduction is about 4.36 deaths per 1,000 births (-18%).

Policy 3: Expanding both access and quality. Ensure all births occur in facilities with a doctor. The three relevant transitions are: facility without doctor $\rightarrow$ facility with doctor for current users, low-risk home $\rightarrow$ facility with doctor, and high-risk home $\rightarrow$ facility with doctor.

Table 1.9. Share of births and neonatal mortality rates by subgroup, defined by risk level, place of delivery, and access to a doctor. Each cell reports the share of births (left) and the neonatal mortality rate (right) for a given subgroup. The risk level is defined based on predicted neonatal mortality from pre-birth characteristics, with high risk indicating observations above the median. Facility with doctor refers to deliveries in facilities for mothers whose closest facility reports at least one doctor on staff at the time of the survey. Neonatal mortality is defined as the probability of death within 28 days of birth.

	Home	Facility without doctor	Facility with doctor
Low risk	4.07% — 0.0437	47.08% — 0.0177	13.43% — 0.0199
High risk	2.42% — 0.0416	27.23% — 0.0324	5.76% — 0.0278

The estimated effect for low-risk births in facilities already with a doctor is zero. The total change is:

$$\Delta_3 = (0.2723 \times (-0.016)) + (0.0407 \times (-0.001)) + (0.0242 \times (-0.017)) = -0.00481$$

$$\text{New average} = 0.02424 - 0.00481 = 0.01943$$

This corresponds to a reduction of about 4.8 per 1,000 births (−19.8%).

Interpretation. The quality margin accounts for nearly all of the potential reduction in neonatal mortality in this setting. Expanding access without improving quality moves the average by a negligible amount, and the combined policy adds little beyond the quality upgrade because the home-birth share is small and its effects on the average are limited.

1.5 Conclusion

Substantial investments in healthcare outreach have greatly expanded access to facility birth in many low-income countries. Yet maternal and neonatal mortality remains high and largely preventable. This paper uses data from Malawi to examine under what conditions accessing delivery care translates into better survival, and how these experiences shape mothers' future engagement with the health system, asking when expanding access saves lives and whether a single encounter can set lasting patterns of care-seeking.

The analysis shows that success depends jointly on who reaches the facility and on the quality they encounter. For mortality, high-quality care delivers large survival gains for high-risk births that have access to a facility where a doctor is available, while accessing facilities yields little benefit and may even worsen outcomes for low-risk births. In terms of future use, exposure to low-quality care reduces subsequent healthcare visits, and the place of delivery for the next birth shows strong persistence regardless of quality.

Taken together, the results suggest that the facility birth rate is not a sufficient policy metric. Lasting progress requires integrating access, quality, and the targeting of high-risk births into a unified strategy. Expanding access without corresponding improvements in quality is unlikely to deliver sustained mortality reductions and may weaken trust in the health system. Policies that push demand toward facilities unprepared to meet it can discourage future use and erode confidence in formal care. Priority should go to skilled staffing and emergency readiness where birth risks are highest. However, deploying highly trained personnel in remote areas is costly, making the optimal degree of skill decentralization a central policy trade-off. Future research should evaluate how to balance centralization of high-skill care with peripheral reach to design sustainable and effective last-mile healthcare delivery.

Chapter 2

Medical Colleges and Front-line Health Workers

2.1 Introduction

Proximity to service providers is essential to access non-tradable services like education and healthcare. Increasing the number of skilled providers in rural health clinics can lead to improvements in health outcomes (Okeke, 2023b; Agte and Soni, n.d.). However, high-skilled teachers and doctors prefer to work in areas with better amenities, leading to a skewed distribution of providers per capita within a country. For instance, the number of doctors per 1,000 people varies fourfold (from 0.5 to 2) across states in India[1] with the disparity being more pronounced across districts and subdistricts. Mandatory rural service has limited tenure and only partly addresses the shortfall. Incentives to work in remote locations are expensive and imperfectly effective.[2]

In this paper we examine an alternative policy option: to set up medical colleges in remote areas and train the locals to serve there themselves. A new college can expand the supply of doctors through two channels: (i) it mechanically raises the number of medical seats available, and (ii) it reduces the distance students must travel to study and lowers the cost of pursuing a medical degree. Because establishing a practice requires local networks and the community's trust, doctors may also find it easier to practice in their home district. We study how opening a medical college in a district affects the number of students from that district who complete a medical degree, how this in turn impacts the number of health facilities and health workers in the district, and how a larger workforce translates into downstream outcomes such as healthcare take-up and health status.

Leveraging the fivefold growth in the number of medical colleges across India between 1980 and 2020, many of which were the first in their district, we answer these questions using a difference-in-differences design. The identifying assumption is that absent the opening of

[1] Data from Indian Medical Register (of India, 2018).

[2] In Odisha, India, the government offers incentives equal to the base salary to attract doctors to remote districts (Hindu, 2015). In China, students who have signed a contract to serve in rural areas have a low willingness to do so (Hu et al., 2023). Even when providers are mandated to serve in rural areas, absenteeism is higher, possibly due to their long commutes from home (Chaudhury et al., 2006).

medical colleges, the districts which received a college earlier and the districts which received a college later would have followed parallel trends in outcomes. This is plausible for two reasons: (i) because it takes many years for a medical college to go from conception to approval to its first graduating batch, the timing at which the first batch graduates is unlikely to respond to short-run outcomes; (ii) both private and public medical colleges are unlikely to be located in response to local trends in the stock of health workers or health outcomes, which are infrequently and poorly measured. Consistent with this assumption, there are no differential pre-trends.

We construct a panel at the district $\times$ year level using multiple datasets. The Indian Medical Register provides the number of residents of each district completing a medical degree each year. Using four rounds of the Economic Census we measure the number of health facilities and the number of employees in these facilities (including doctors, nurses, and other personnel). Four rounds of the Demographic and Health Surveys (DHS) and three rounds of the District Level Household and Facility Survey (DLHS) measure the take-up of care by pregnant women and health outcomes for infants and adults. Data from the National Medical Commission give us the location and year of opening of every medical college in India. The treatment effects are estimated using a two-way fixed effects specification with treatment defined as the number of cohorts that have graduated from a medical college in that district. We control for the presence of a college in the district, which captures the effect of opening a college prior to any batch graduating, and for spillovers from colleges in nearby districts. All results are robust to estimating using the specifications in Callaway and Sant'Anna (2021).

The opening of a medical college leads 16 additional students from that district to complete a medical degree, relative to other districts in the same state. With a baseline of 10.7 students per district per year, this constitutes a large 150% increase. Together with the mechanical impact of increasing the number of medical seats, this represents a substantial expansion in the supply of doctors from that district. We find that one additional batch of students graduating from a medical college in a district leads to 31.5 additional health facilities in the same district. This amounts to a 4.3% increase each year relative to a mean of 722 facilities in districts where colleges had not yet opened in 1990. This effect is in addition to a one-time increase of 200 facilities when the college opens. The number of employees at health facilities increases by 145 with each graduating batch[3]. The increase is driven by those providing western allopathic medicine, with no effect on the number providing alternative systems of medicine. This is consistent with colleges only training students in western medicine. We find no impact on the number of public facilities or public providers; the increase is concentrated in the private sector and in urban areas. The opening of a medical college hence substantially increases access to providers in that district.

The increased access, however, leads to only modest improvements in the take-up of healthcare: there is no significant increase in the number of antenatal care (ANC) visits by pregnant women although there is a slight increase in the share of ANC visits attended by a doctor. Similarly, we find a slight increase in the share of deliveries assisted by a doctor but an overall decrease in the share of deliveries assisted by skilled personnel, suggesting substitution toward doctors rather than a net gain in skilled attendance. There is a small and statistically significant decrease in the incidence of diarrhea, cough, and fever among children but a small and statistically insignificant increase in the infant mortality rate. Taken together, these effects are economically negligible. Improving health outcomes may therefore require complementary policies: hiring the additional doctors into the public system, posting them to rural health centers, and providing them with adequate infrastructure.

This paper contributes to the literature on the distribution of high-skilled workers like doctors. Better amenities in urban areas and agglomeration benefits influence high-skilled workers' preference to be in cities (Arntz et al., 2023; Diamond, 2016; Jeworrek and Brachert, 2022). This is a problem in the case of non-tradable services like healthcare and has been long discussed (Newhouse et al., 1982; Rosenthal et al., 2005). Dingel et al. (n.d.) suggests there may be increasing returns to scale in the provision of advanced care and hence such a distribution may be efficient. However, it is desirable to make primary care accessible to all. Inducing doctors to

[3]Each batch consists of 100 students on average.

serve in rural areas through incentives and mandates are only partly effective. The required wage differential is very large, making this expensive (Holte et al., 2015; Scott et al., 2013; Swami and Scott, 2021). In this paper, we explore a second policy option: we show that opening colleges in remote areas could lead to an increase in health workers in those areas. This takes advantage of potential heterogeneous location preferences in the population, as students studying in remote colleges may be more willing to serve there.

This paper also relates to the literature on the local impacts of higher educational institutions. Across countries, when a college opens nearby, studies find improvements in college enrollment among students and, in turn, labor market returns (Frenette, 2006, 2009; Lapid, 2018). Jagnani and Khanna (2020) also see more enrollment in primary and secondary schools and a crowd in of services like electricity and roads. Large university expansions in China and Vietnam have also been found to have lead to an increase in firm productivity, capital investments in firms, and structural transformation towards high-skill industries in the neighbourhood. This leads to greater returns for the educated (Vu and Vu-Thanh, 2024; Che and Zhang, 2018). In this paper we show that setting up medical colleges has large externalities on access to healthcare locally. These improvements can likely benefit all people in those districts and need to be factored in to the returns to starting such colleges.

This paper also contributes to an active policy debate on what governments can do to increase the concentration of providers in remote areas. The government in South Korea is increasing the number of medical seats in existing colleges hoping to achieve this while doctors are protesting this move claiming it would only lead to more doctors in the urban areas (Pacific, 2024). In India, the National Medical Commission has proposed that new medical colleges be opened only in states with few colleges[4]. Muralidharan (2024) suggests that hiring public workers locally may make them more regular to their workplace and be accountable to the population they serve. In this paper, we show that this is possible: opening higher educational

[4]The move was intended to balance the distribution of medical colleges across the country. The proposal is to cap the number of seats to 1 per 10,000 population.

institutions does lead to more providers in that district. Reforming the public hiring process, like creating district-level cadres will allow the state to capitalize on its presence in the district.

2.2 Context: Medical Education in India

After high school, students who want to pursue a medical career can enroll in a Bachelor in Medicine and Bachelor in Surgery (commonly referred to by the abbreviation MBBS). The usual duration of an MBBS degree is five and half years and includes a year of residential internship. After completing this degree the student is eligible to practice allopathic medicine. There are degrees in other forms of medicine like Ayurveda, Unani, Siddha, and homeopathy. These courses are usually offered in other colleges that specialize in these schools of medicine.

2.2.1 Medical colleges

There are about 700 medical colleges in India as of 2023. The initial medical colleges were started by the state or central government. Subsequently non-profit trusts and charitable societies were allowed to set up colleges. The principle that medical education cannot be a profit-making venture prevented private bodies from starting colleges until this was relaxed in 2017 (Mahal and Mohanan, 2006).

The process of setting up a private medical college is lengthy and cumbersome. The college needs to be attached to an existing public or private hospital where students would receive practical training. The hospital needs to have at least 300 beds and the capacity to increase number of beds based on the number of students enrolled. The private entities need to own a land at least 20 acres big where the college can be set up (10 acres for big cities). They need to have the required infrastructure for the college, the attached teaching hospital and residential quarters for students. They also need to provide bank guarantees to establish their ability to run a college. They first need to get an approval from the university they wish to be affiliated with and the state government where the college is located. They can then submit the request to the central government. The request is forwarded to the Medical Council of India (now renamed

National Medical Commission) who conduct the technical scrutiny of the application and request clarifications from the applicant. Once all requirements are met the private entity finally receives the approval to start the college. The entire process could take multiple years.

2.2.2 Medical admissions

Process of admissions to medical colleges was determined by individual state governments. Starting in the 1980s, they conducted a common entrance exam for the colleges in the state. Students scoring above the cutoff for a college were granted admission to that college. 80% of the seats were reserved for the domiciles of that state. While domicile rules vary across states, usually the student should have completed a few years of education in the state or one of their parents should be a resident of the state for them to be considered a domicile. In states with multiple medical colleges, the students get to choose the college they study and proximity to their homes is one of the important factors in this choice. Since 2016, a centralized exam has been conducted across the country to determine admissions to medical colleges. However, the domicile restrictions were retained as before. 15% of the seats in private colleges are also reserved under a management quota. Admissions to these seats are at the discretion of the college administration.

Fees to pursue a medical degree in public colleges are relatively low (about 500 USD per year), while fees in private colleges are very high (about 30,000 USD per year). In addition to the fees, students' costs include living expenses near their college or commuting costs to and from the college. Hence, students might choose to study in colleges closer to their homes to reduce these costs.

2.2.3 Medical jobs in India

After completing their MBBS degrees, students may pursue a post-graduation degree or other specialized degrees. The number of post-graduation seats in the country is less than half of the number of undergraduate seats. Admissions to the post-graduate seats are based on another entrance test conducted nationally.

Public recruitment of doctors usually happens at the state level. State governments conduct exams to award jobs. Doctors who are awarded the job are added to the state public service cadre, and they may be assigned to any location in the state. Public doctors are also often transferred between different locations.

With limited public recruitment, many doctors work in private hospitals and clinics, or set up their own private practice. Jobs in the private sector are usually obtained through connections the student can leverage besides their performance in their degree. Trust among their potential patients is essential for a doctor to set up a private practice. This makes it difficult for a doctor to go to a new city and set up a clinic. Hence they are more likely to return to their home district to start a clinic or work in a hospital.

2.3 Data

Colleges: We use the list of medical colleges approved by the National Medical Council. This public database contains the address of the college, the year it received the approval, the number of seats, and the administration type (public, private, or trust). We assume that the year of approval is the year the first batch of students were admitted to that college. We also assume that no medical colleges have closed down in this period. We assign the colleges to the district where they are located in 2011.

The number of medical colleges increases from 112 in 1980 to 555 in 2020. Of the 112 colleges in 1980, 99 were government-run colleges and 13 were privately managed. Of the 555 colleges in 2020, 275 were public colleges and 280 were privately managed. Hence, of the 443 colleges that opened between 1980 and 2020, 40% are government-run and 60% are privately managed.

Figure 2.2 presents the distribution of these colleges across the country in 1980 and 2020. While in 1980 there are colleges across the country, most of them are located in the state capitals

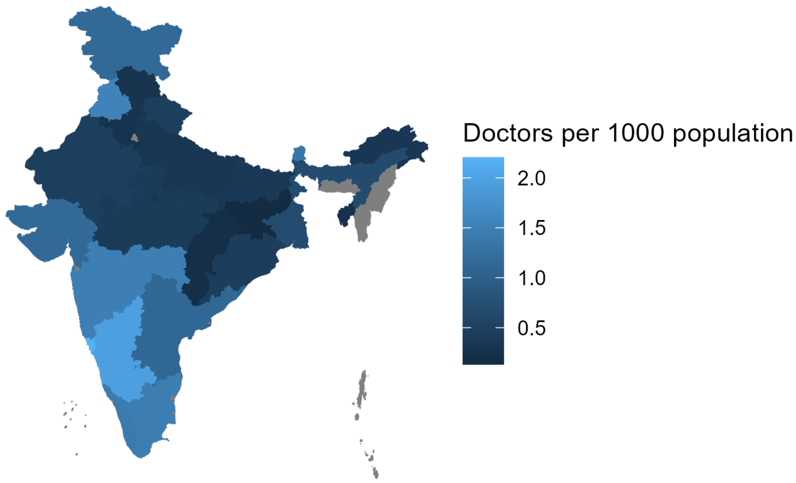
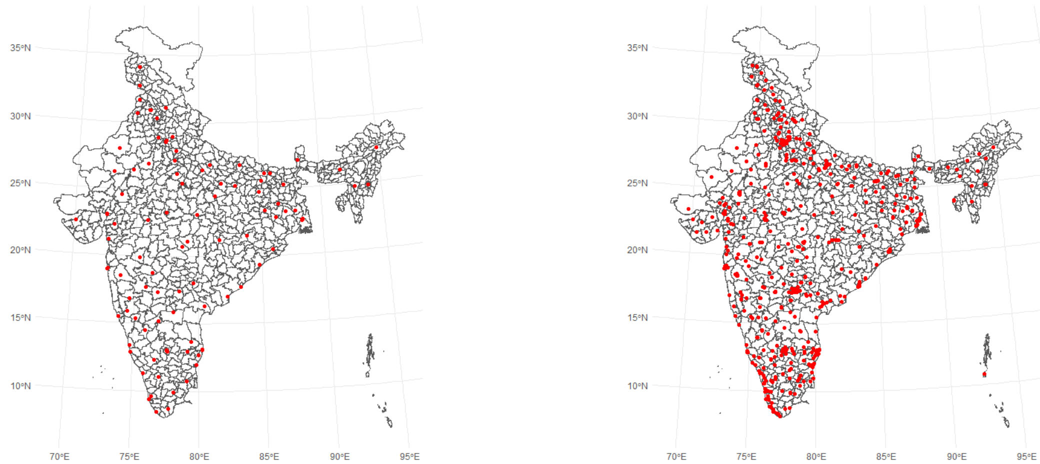


Figure 2.1. Doctors per 1000 population across states.

and in the bigger cities with few colleges in smaller cities and remote districts. The number of districts with a medical college increased from 89 in 1980 to 294 in 2020.



(a) 1980 **(b)** 2020

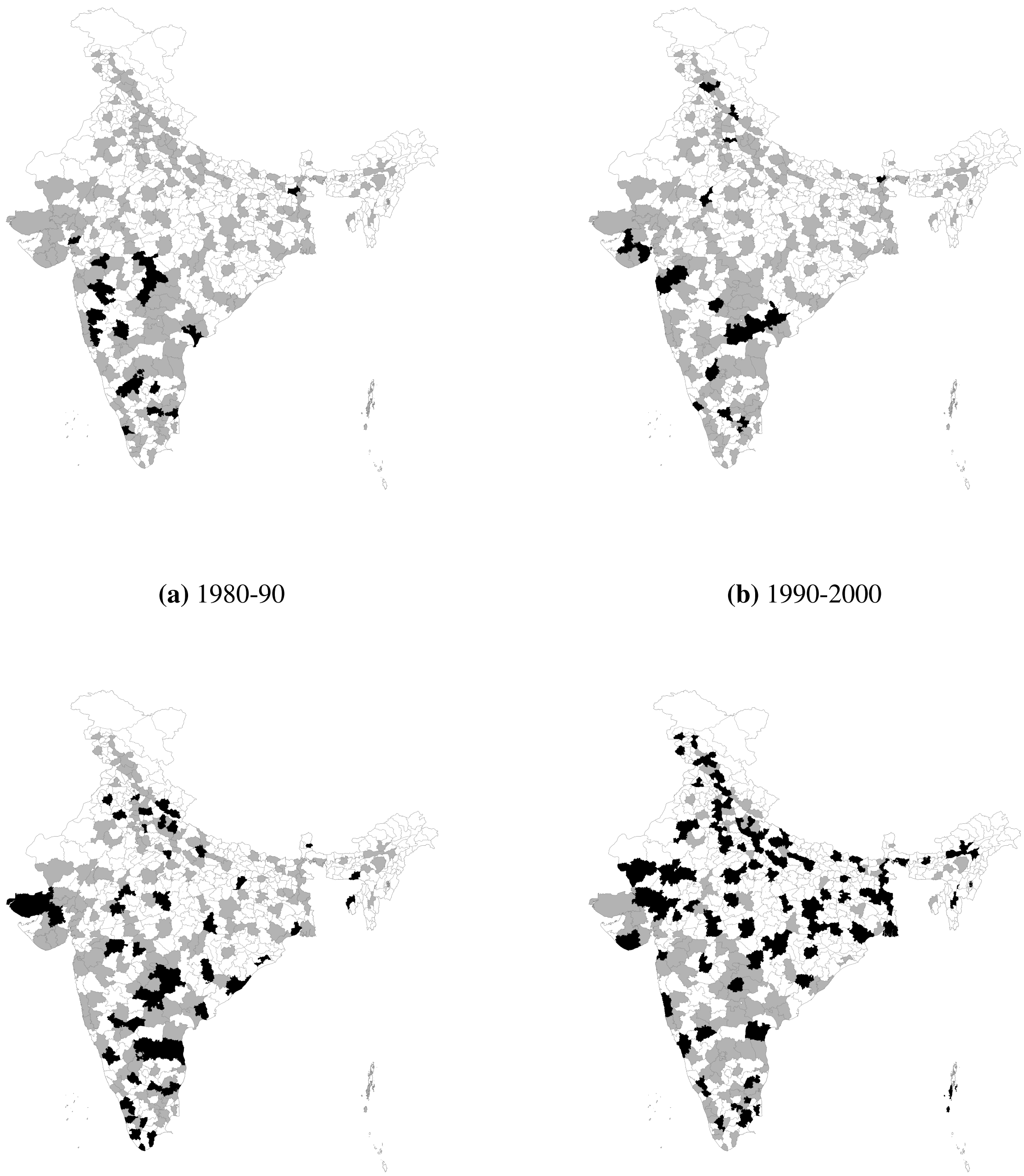
Figure 2.2. Location of medical colleges across the country.

Health facilities and workers: We use four rounds of the Economic Census to compile outcome measures. The Economic Census is the complete count of all entrepreneurial units located in the country involved in any economic activities. We use the three-digit NIC codes of each establishment to count the number providing health services. This includes all organizations providing health and medical services like hospitals, dispensaries, sanatoria, nursing homes, and maternal and child welfare clinics. We further use the four-digit codes to classify these institutions by the type of medicine practiced: allopathic, Ayurvedic, Unani, homeopathy, and others. We group the last four categories as non-allopathic for convenience.

The census data also includes the number of employees in each facility. We do not observe any characteristics of these employees other than their gender. The employees may be doctors, nurses, pharmacists, accountants, sanitary staff, or any others who work in these facilities. We argue that the total number of workers in the facility is a reasonable proxy for the number of health workers in the district. We aggregate all census data up to the district level.

Healthcare take-up and outcomes: The DHS is a nationally representative household survey that provides data on population, health, and nutrition in developing countries[5]. We use

[5]The DHS is called NFHS in India and is administered by an independent organization called Indian Institute of Population Sciences.



(a) 1980-90 **(b)** 1990-2000

(c) 2000-10 **(d)** 2010-20

Figure 2.3. Districts by year of first opening of medical colleges.

four rounds of the DHS to study outcomes related to healthcare take-up and health status. These are the first, second, fourth, and fifth rounds administered in 1992-93, 1998-99, 2015-16, and 2019-21 respectively. We do not include data from the third round since it does not contain district identifiers. Along with the DHS, we also include health outcomes data from the District Level Household Surveys (DLHS) datasets[6]. We use three rounds of the DLHS, viz. DLHS-2, DLHS-3 and DLHS-4 spanning interview years 2002-04, 2007-08 and 2012-14. The DLHS is yet another nationally representative survey, originally designed to measure reproductive and child health (RCH) outcomes. We do not incorporate the DLHS-1, since mapping the district codes requires a long digitization process. We use these seven datasets together at the individual birth (child) level for birth/pregnancy/antenatal care outcomes. The reference year for pregnancy/birth outcomes is considered as the birth year itself. The recall period for certain specific variables such as whether a child has had diarrhoea, fever, or cough is just two weeks before the survey, so the interview year is considered as the reference year for these particular outcomes. Outcome variables are harmonized across all datasets to maintain comparability. DHS and DLHS women weights are used, which are normalized across the datasets. To estimate the neonatal, infant and under-5 mortality rate, binary variables are created to tag each birth whether a child was reported dead within one month, one year and five years of its birth. Moreover, delivery providers and antenatal care outcomes are only present in the DLHS for the latest birth.

Medical doctors registrations. We also incorporate data from the Indian Medical Register, which contains comprehensive information about the universe of ~ 1.3M doctors registered in India from 1960 to 2020.[7] For each doctor, the register specifies their home district and pin code at the time of registering, the university in which they studied, the year of graduation, the year of registration, and the state medical council they are registered in.

We assume that the district and pin code at the time of registration corresponds to their permanent address or their parent's address. Students typically register at the end of their

[6]The DLHS was also administered by Indian Institute of Population Sciences.
[7]The register is incomplete post 2020.

education (under graduation or post graduation) and before starting a job. They typically live in a college hostel with a temporary address. Their permanent address would correspond to their parent's address. If the parents have not migrated in the last 6 years, this is also the district they would have been living in before starting their degree. We provide multiple tests of this assumption and fail to reject it. Anecdotal evidence from multiple doctors also confirms this trend. We also restrict the analysis to doctors who registered in the same year as graduating from under graduate degree for whom the assumption is more likely to hold.

2.4 Econometric Strategy

We are interested in studying the impact of opening medical colleges in a district on the number of health workers in the same district. We capitalize on the staggered opening of medical colleges across the country to estimate the treatment effects using a two-way fixed effect specification. The unit of analysis is district x year. In Figure 2.3 we see the staggered opening of the first college in a district across four decades.

In the TWFE specification, we account for the fact that the number of health workers is a stock variable that accumulates over time post-treatment. We also acknowledge that the effects are not limited to the same district and likely have a spillover effect on neighboring districts within the same state.

We estimate the parameters in the following specification:

$$y_{idt} = \theta_t + \eta_d + \beta batches_{dt} + collegeopen_{dt} + \gamma spillovers_{dt} + \varepsilon_{idt} \qquad (2.1)$$

where y_{idt} is the outcome of interest for birth i in district d in year t, θ_t are year fixed effects, and η_d are district fixed effects. $batches_{dt}$ is the treatment variable that is equal to the number of batches graduated from a college in the mother's district of residence. We assume that the treatment effects aggregate over time linearly here. In alternative specifications we also estimate the heterogeneous impact over time non-parametrically using indicators for 0-5 years, 5-10 years, 10-15 years and >15 years since treatment. $collegeopen_{dt}$ is an indicator that is 1 if there is a college in the district. This variable captures the time-invariant average effect of opening a college. Although this term is also a treatment effect, we do not emphasize this in the results since it is a more mechanical effect of opening a large institution in the district and less interesting economically.

For outcomes measured at the district level (e.g. number of health workers, number of health facilities) we adapt this specification to district-level data. The unit of observation becomes the district, and the treatment variable remains defined at the district-birth-year level. For medical doctor registrations, we include state-year fixed effects to account for the change in medical college seats capacity within the state.

$spillovers_{dt}$ is a variable that captures the total spillovers of a college opening on all other districts. This is defined as the number of batches graduated in colleges in other districts weighted by the inverse of the distance from that college. It is defined as:

$$\text{spillovers}_{dt} = \sum_{\text{colleges } c} \frac{\mathbb{1}[state(c)=state(d) \;\&\; district(c)\neq d] * \mathbb{1}[opening(c)<t] * (t-opening(c))}{(distance(district(c),d))^2}$$

Here, we make three assumptions about the spillovers: (i) the spillovers are restricted to the districts in the same state of the college and decay with the square of the distance from the college; (ii) the spillover effects aggregate linearly over time (or are proportional to the number of batches of students that have graduated); (iii) the total spillovers on a district is the sum of spillovers from all colleges in the state. In alternative specifications we vary these assumptions in the spillover definition.

We exclude districts that are always treated (have a college in the first time period) and never treated (don't have a college even in 2020). Besides the above functional form assumptions, we also assume parallel trends between counterfactual outcomes in the districts with a college and the observed outcome in the districts without a college as yet conditional on the spillover and fixed effects.

We test this assumption by comparing trends for districts where colleges opened earlier with districts where colleges opened later. We focus on the subset of colleges that opened between 2000 and 2020. We compare districts where a college first opened between 2000 and 2010 with districts where a college first opened between 2010 and 2020. We compare the outcomes of these districts in the previous periods using the following specification:

$$y_{dt} = \theta_t + \eta c_d^{2000-10} + \beta_t c_d^{2000-10} + \nu_{dt} \tag{2.2}$$

where $c_d^{2000-10}$ is an indicator that is 1 if a college first opened in that district in 2000-10 and 0 otherwise.

We claim that under these assumptions β is the treatment effect of opening a medical college in that district. It is the difference in potential outcomes of that district if that particular college was not opened versus if that college was opened, holding all other college openings the same.

The interpretation of the spillovers term is however difficult. The districts receiving a higher value of the spillovers may not have parallel trends with the districts receiving a lower value of the spillovers. For example, the districts receiving a higher spillover may be those near bigger cities where many colleges open in this period. The districts closer to the cities may not be comparable with the districts further away. Hence, we do not analyse the coefficients of the spillover term in this paper. We include them only to remove the bias in the treatment effect due to the spillovers.

Robustness: Callaway and Sant'Anna estimator

To ensure the results are robust to the specification, we also use the estimation procedure proposed by Callaway and Sant'Anna (2021). We define the dependent variable as an indicator of whether there is a medical college in the district in that year. We estimate the heterogeneous treatment effects over time. Consider the data generating process given by:

$$y_{dt} = \theta_t + \eta_d + \sum_{\ell} \beta_\ell \mathbf{1}\{t - g_d = \ell\} + v_{dt}$$

where y_{dt} is the outcome of interest in district d in year t, θ_t are year fixed-effects, η_d are district fixed effects, g_d is the year in which district d had the first medical college opening. The β_ℓ parameters are the average treatment effect ℓ years from the college opening.

We again use the not-yet-treated districts as our comparison group. For the Economic Census outcomes, we define a time period as 8 years long. Hence a college may have opened in any of those 8 years. We estimate the average treatment effects in the next period in all such districts.

2.5 Results

We first examine the impact of college opening on the number of individuals from the same district who enroll for MBBS and complete the degree. These estimates compare the numbers across districts in the same state to account for the mechanical effect of an increase in the number of medical seats. The event study estimates in Figure 2.4 reveal a clear positive effect in the years following the opening. The estimates are near zero in the pre-opening period, confirming parallel trends. By years 5–15 post-opening, the average treatment effect on the treated (ATT) reaches about 15 registrations. Relative to a baseline 10.7 registrations per year per district in districts without a medical college yet, this constitutes a 140% increase which is substantial. Thus the opening of a medical college leads to a disproportionate number of students from the same district enrolling in and completing medical degrees.

We then look at the impacts on the number of health facilities and number of health workers. In Table 2.1 we see the results on the number of health facilities in the district. One additional batch of students graduating from a college in the district is associated with 31.5 additional health facilities in the district. This is a large effect relative to the mean of 722.1 facilities in 1990 in districts with no medical colleges yet. We see that most of these additional

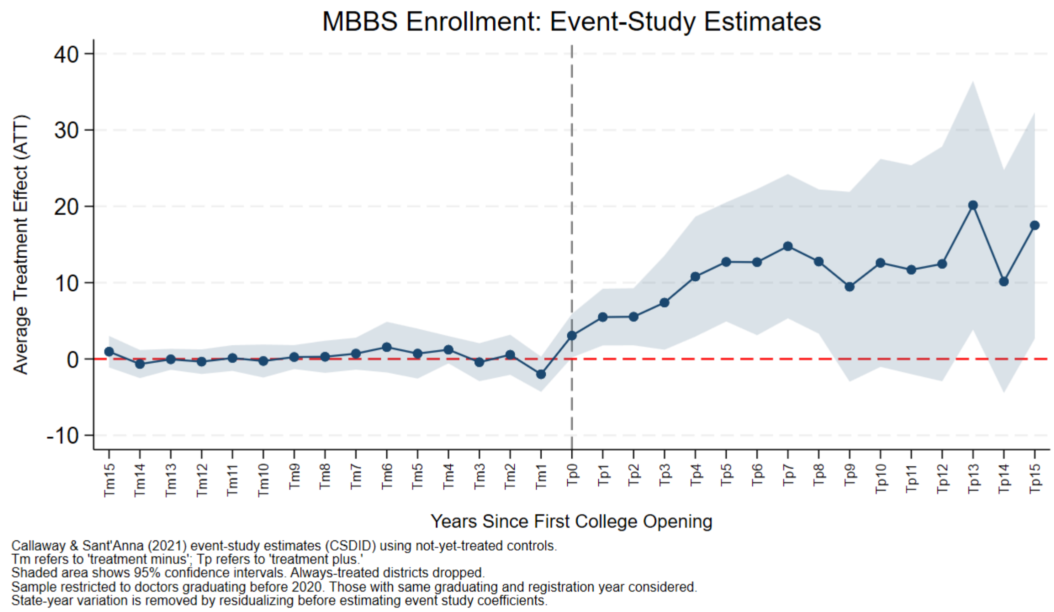


Figure 2.4. Effect on local doctors registrations. CSDID estimates of doctors joining (Years post 2020 dropped from registry data. Graduates with same registration and graduating year. Residualized outcome variable to remove state-year variation before running csdid), (-15, +15) periods.

facilities are private facilities, practicing allopathic medicine in urban areas. We see an increase of 26.9 urban health facilities, 31.4 private health facilities, and 33.1 allopathic health facilities with each batch of graduating students. There is a direct effect of opening a college as well. We see an average 199.6 additional facilities in the district before any batches graduate.

In Table 2.2 we see the treatment effects on the number of people working in health facilities in the district. This is the sum of the number of employees working in all establishments classified as health facilities in the district and may include non-medical staff working there too. One additional batch of students graduating from a medical college in the district is associated with 144.7 additional employees at health facilities. 74.7 of these workers are male and 70 are female. The additional workers are mostly in the private sector (138.4) and in urban areas (141.4).

We look at health services take-up and health outcomes from DHS and DLHS surveys in India. In Table 2.3, we see no significant change in the number or timing of ANC visits received

Table 2.1. Impact of colleges on health facilities. This table presents the results on the number of health facilities and by type of facilities. The coefficients estimated are the parameters in equation 2.1. All regressions include district and time fixed effects.

	Dep. Var. : Number of Health Facilities						
	(1) All	(2) Rural	(3) Urban	(4) Public	(5) Private	(6) Allopathic	(7) Non-Allop.
Graduate batches	31.5***	4.6*	26.9***	0.3	31.4***	33.1***	-1.6
	(4.2)	(2.7)	(2.6)	(1.3)	(3.8)	(3.8)	(2.8)
College opening	199.6***	69.5***	130.1***	22.8**	174.9***	127.0***	72.5***
	(30.4)	(20.0)	(19.1)	(9.3)	(27.2)	(27.3)	(20.3)
Spillover	37.1***	4.1	33.0***	-5.0*	42.1***	29.3***	7.8
	(9.1)	(6.0)	(5.7)	(2.8)	(8.2)	(8.2)	(6.1)
Observations	2,054	2,054	2,054	2,054	2,054	2,054	2,054
Mean Dep. Var.	722.1	431.6	290.5	126.3	595.2	551.9	170.3

Table 2.2. Impact of colleges on workers in health facilities. This table presents the results on the number of workers employed in the health facilities. The coefficients estimated are the parameters in equation 2.1. All regressions include district and time fixed effects.

	Dep. Var. : Number of Health Workers						
	(1) All	(2) Female	(3) Male	(4) Rural	(5) Urban	(6) Public	(7) Private
Graduate batches	144.7***	70.0***	74.7***	3.2	141.4***	6.3	138.4***
	(24.2)	(8.7)	(20.2)	(19.5)	(13.7)	(5.4)	(23.2)
College opening	1288.5***	762.5***	526.0***	326.3**	962.2***	134.2***	1154.2***
	(175.9)	(63.0)	(146.3)	(141.4)	(99.3)	(38.9)	(168.6)
Spillover	167.4***	66.4***	101.0**	32.6	134.8***	-9.4	176.9***
	(52.7)	(18.9)	(43.8)	(42.4)	(29.8)	(11.7)	(50.5)
Observations	2,054	2,054	2,054	2,054	2,054	2,054	2,054
Mean Dep. Var.	2159.0	661.6	1497.3	1025.5	1133.4	756.6	1402.4

Table 2.3. Impact of colleges on take-up of healthcare. * p<0.10, ** p<0.05, *** p<0.01. Standard errors clustered at the district level. District as well as survey-year fixed effects used. The coefficients estimated are the parameters in equation 2.1. Variables are available in all datasets, viz. 4 rounds of DHS and 3 rounds of DLHS. 'Any ANC' variable is reported as a percentage. 'ANC visits' denotes the number of visits, 'First ANC' denotes the first month of ANC, for those who had an ANC visit. Variables available for the last pregnancy.

	Dep. Var.: Health Services Uptake		
	ANC Visits	First ANC	Any ANC
	(1)	(2)	(3)
Graduate batches	0.007	-0.002	0.006
	(0.004)	(0.003)	(0.009)
College opening	0.068	0.024	0.152
	(0.050)	(0.029)	(0.167)
Spillover	0.062	0.064***	-0.146
	(0.051)	(0.020)	(0.105)
Mean Dep. Var.	3.682	2.817	67.387
Observations	769575	705765	3435640

by pregnant women. In Table 2.5 we see a small but significant increase in the share of ANC visits attended by doctors of about 0.17pp and a decrease in the share attended by nurses, TBAs, anganwadi and ASHAs. This suggests doctors are substituting nurses and other workers.

In Table 2.4, we find a 0.30 percentage point decrease in births assisted by skilled personnel. This reflects a compositional shift: deliveries by doctors increase by 0.17 percentage points while those by nurses fall by 0.43 percentage points. Institutional childbirth rates are unchanged.

In Table A.2.1 we see that the increase in ANC visits is mostly urban areas. However, the substitution of ANC visits and facility deliveries conducted by doctors instead of nurses takes place mainly in rural areas, as shown respectively in Table A.2.2 and A.2.3.

In the first four columns of Table 2.6 we see a small but significant decline in the incidence of diseases like diarrhea, cough and fever. In the last three columns we see the effects on early-life mortality. We rule out even a small decline in neonatal and infant mortality rate. The point estimates for under 5 mortality are positive and significant, but of a very small magnitude

Table 2.4. Impact of colleges on share of deliveries assisted by provider type. * p<0.10, ** p<0.05, *** p<0.01. Standard errors clustered at the district level. District as well as survey-year fixed effects used. The coefficients estimated are the parameters in equation 2.1. All variables are reported as percentages. Variables are available for women who have had at least one birth. Delivery provider variables are defined for the last birth in DLHS and for available births in DHS. Delivery by doctor, nurse, TBA, none, or other are available in all datasets. Skilled birth denotes if the birth was assisted by either a doctor, nurse or a trained health professional. Here, TBA means 'traditional birth attendant'. The category of 'other health professional' not reported here since its mean share is merely 1.7%. Facility denotes whether the birth took place in any medical facility, and not at home/on the way to a facility.

	Dep. Var. : Share of Deliveries Assisted						
	Facility	Skilled Birth	Doctor	Nurse	TBA	Other	None
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Graduate batches	-0.010	-0.296***	0.173**	-0.428***	0.147*	-0.020	0.009
	(0.072)	(0.067)	(0.086)	(0.108)	(0.077)	(0.022)	(0.009)
College opening	0.647	-2.135***	1.776**	-2.907***	-0.091	0.061	0.019
	(0.639)	(0.737)	(0.898)	(0.902)	(0.788)	(0.180)	(0.067)
Spillover	-0.514	-2.229***	-0.194	-2.953***	-0.956	0.949***	0.181***
	(0.469)	(0.611)	(0.569)	(0.892)	(0.677)	(0.208)	(0.060)
Mean Dep. Var.	48.120	54.967	30.063	35.061	25.402	5.256	0.635
Observations	1166496	751797	842069	842069	842069	827374	842069

Table 2.5. Impact of colleges on share of ANC visits by provider type. * p<0.10, ** p<0.05, *** p<0.01. Standard errors clustered at the district level. District as well as survey-year fixed effects used. The coefficients estimated are the parameters in equation 2.1. All variables are reported as percentages. ANC_none is available in all datasets. ANC providers are present in DHS datasets and DLHS 2003. Within these, Anganwadi and ASHA are only present in DHS. ANC provider questions are asked to those who saw someone for an ANC. Categories of 'village health worker' and 'other' ANC provider not reported here, since their mean shares are merely 0.63% and 0.19% respectively.

	Dep. Var.: Share of Antenatal Care Visits					
	Doctor	Nurse	ASHA	TBA	Anganwadi	None
	(1)	(2)	(3)	(4)	(5)	(6)
Graduate batches	0.171*	-0.149**	-0.711***	-0.009*	-0.398***	0.124**
	(0.099)	(0.059)	(0.107)	(0.005)	(0.143)	(0.050)
College opening	3.172***	0.200	-0.254	-0.094	-0.445	0.219
	(0.868)	(0.888)	(0.673)	(0.068)	(0.733)	(0.625)
Spillover	0.312	-1.206**	-1.921	-0.012	-0.789	-0.547
	(0.496)	(0.547)	(1.356)	(0.038)	(1.450)	(0.395)
Mean Dep. Var.	36.950	26.146	12.540	0.752	14.866	17.571
Observations	795434	794802	327165	508438	327165	1029918

Table 2.6. Impact of colleges on health status. * p<0.10, ** p<0.05, *** p<0.01. Standard errors clustered at the district level. District as well as survey-year fixed effects used. The coefficients estimated are the parameters in equation 2.1. Morbidity variables reported as percentages, mortality variables reported per thousand. Mortality variables available in all datasets, viz. four rounds of DHS and three rounds of DLHS. Restricted to 5 years before the interview year for mortality variables. Morbidity variables only in DHS, with reference year as the current interview year. Given that only one kind of data (i.e., only DHS) is used for morbidity, dataset fixed effects not used. Neonatal, infant and under-5 mortality refer to death within one month, one year and 5 years of birth, respectively.

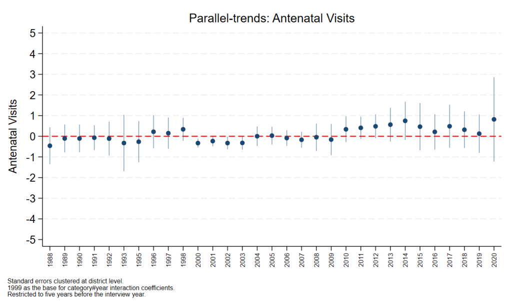
	Child Morbidity			*Child Mortality*		
	Diarrhoea	Fever	Cough	Neonatal	Infant	Under-5
	(1)	(2)	(3)	(4)	(5)	(6)
Graduate batches	-0.139**	-0.147**	-0.170**	-0.066	0.165	0.308**
	(0.057)	(0.066)	(0.074)	(0.122)	(0.134)	(0.138)
College opening	0.075	0.337	0.431	-0.427	0.685	2.058
	(0.638)	(0.819)	(0.706)	(1.019)	(1.431)	(1.658)
Spillover	-0.655**	-0.588	-0.648*	0.828	1.382	2.626***
	(0.264)	(0.375)	(0.387)	(0.733)	(0.874)	(0.974)
Mean Dep. Var.	8.823	14.248	13.983	29.618	42.631	49.405
Observations	404437	404566	404530	1089152	1089152	1089152

(0.031pp).

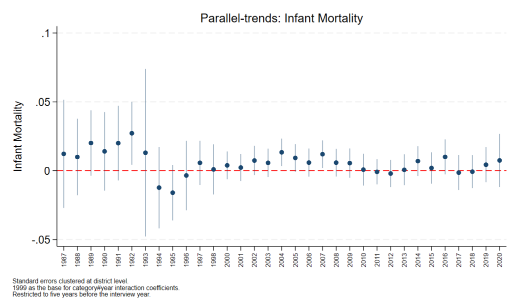
Parallel pre-trends: In Table 2.7 we look at the pre-trends between the districts where a college first opened between 2000 and 2010 and districts where a college first opened between 2010 and 2020. These coefficients are estimated according to the specification in equation 2.2. We see that the coefficients in the pre-period of 1998 are small and insignificant across all variables. The coefficients in 2005 are also small and insignificant, despite some districts may have had a college opened by then. In Figure 2.5 we see the same coefficients for the DHS outcome variables as well. Here again we see that the coefficients are small and statistically insignificant.

Table 2.7. Pre-trends in outcome variables. This table presents the trends in outcome variables between districts where a college first opened in 2000-10 and districts where a college first opened in 2010-20. Each column shows the trend in a different outcome variable. The estimates are of the parameters β_τ in Equation 2.2.

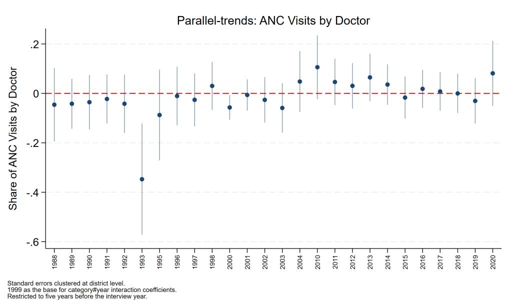
	Test for Parallel pre-Trends					
	(1) Health Workers	(2) Health Facilities	(3) HF: Urban	(4) HF: Rural	(5) HF: Public	(6) HF: Private
1990	0.0	0.0	0.0	0.0	0.0	0.0
	(.)	(.)	(.)	(.)	(.)	(.)
1998	-275.3	-146.8	-8.3	-138.5	-49.0	-97.8
	(539.4)	(215.5)	(89.1)	(154.4)	(34.1)	(203.4)
2005	367.9	65.4	24.1	41.4	-31.9	94.1
	(538.7)	(215.2)	(89.0)	(154.2)	(34.1)	(203.1)
2013	1791.5***	97.5	185.6**	-88.1	-18.5	116.0
	(538.7)	(215.2)	(89.0)	(154.2)	(34.1)	(203.1)
Observations	628	628	628	628	628	628



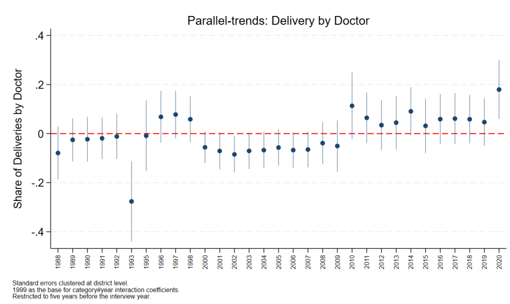
(a) ANC visits



(b) Infant mortality rate



(c) Share of ANC visits by doctors



(d) Share of deliveries assisted by doctors

Figure 2.5. Pre-trends in outcome variables. These figures present the trends in outcome variables between districts where a college first opened in 2000-10 and districts where a college first opened in 2010-20. The estimates plotted are the coefficients β_τ in Equation 2.2.

2.6 Discussion

In this paper, we find that opening a medical college leads to a significant increase in the number of health facilities and health workers in the area. This in turn leads to an increase in the take-up of healthcare like ANC visits. However we reject even small changes in health outcomes like infant mortality rate and incidence of fever and diarrhoea.

The census data does not allow us to distinguish between health clinics run by quacks and those managed by qualified doctors. However, we do see more ANC visits conducted by doctors. We also see an increase in facilities providing western medicine and not alternative systems of medicine. Both of these suggest the additional health workers in the district are not all quacks. An increase in the number of qualified doctors could lead to a substitution of care sought from quacks with care sought from doctors. Further, Das et al. (2022) show that the quality of care provided by quacks is better in places where quality of care provided by qualified providers is better, perhaps driven by competition. Hence even if the increase in number of doctors is only a fraction of the observed treatment effects, there can be a large increase in access to quality care.

In this version of the paper, we are unable to identify the mechanisms at play here. Students who live near these colleges may choose to attend them and then practice near their home districts. Students from across the state may come to the college and choose to stay in the same district for work. We will be able to separate these mechanisms using additional data available with the government.

Reducing disparities in the distribution of health providers is of considerable policy interest in both developing and developed countries. Government of South Korea announced in 2024 that they would be increasing the number of medical seats to have sufficient doctors serving in the remote areas of the country. In India, a new regulation restricts opening of new medical colleges to states with a doctor per 1000 people ratio below 1.

However, disparities in the distribution exist at multiple levels. In India, in figure 2.1 we see the disparities across states. Within each state though, the districts containing big cities may have more doctors. Further within each district, the doctors may live in more urban areas. While the opening of colleges in remote districts can reduce the former two imbalances, the last one persists. We see that the increase in providers and facilities is concentrated in the urban areas of the district. While people in the rural areas can travel about 15-20 kms to reach the urban areas, the severely ill may not be able to do so.

Improving health outcomes is very difficult. The marginal patients seeking care may not be the severely ill. Further, Weaver et al. (2024) show that there are complementarities between factors like healthcare, nutrition, and sanitation in how they affect outcomes. Hence improving any one factor alone may not move the needle significantly.

More specifically, there may be complementarities between provider knowledge and the equipment and infrastructure at the facilities. While we see the largest increase in *private* providers and facilities, the *public* facilities are better funded and may have better equipment. Assigning the additional providers to public facilities in rural areas may hence be more effective.

One reason this may not be happening currently is due to the structure of public hiring in India. State governments hire doctors at the state level and may assign them to any facility in any district in the state. This mechanism does not take advantage of the doctors' preferences to be in some districts than others. As a result most doctors vie for postings in the state capital. The state could establish district-level cadres of doctors, recruiting and appointing them within the same district. This approach might offer significant benefits for the medical professionals, as it provides the certainty of employment within a single district. Consequently, doctors would not need to relocate their families and could confidently invest in their local communities. In this paper we show that colleges expand the pool of candidates available in the district. The increased availability of qualified doctors directly supports the feasibility and success of a strategy of local recruitment and retention of health professionals.

Acknowledgments. Chapter 2, in full, is currently being prepared for submission for publication of the material. Budillon, Fulvia; Ramachandran, Sabareesh. The dissertation author was one of the primary investigators and authors of this material.

Chapter 3

Attribution and Learning from Local Health Outcomes

3.1 Introduction

People often must make decisions based on outcomes they observe in their community: the failure of a business in their town, the outcome of a peer's investment, the success of a treatment they see in a neighbor. These outcomes are informative but rarely transparent. They reflect the quality of the choice that produced them, but also the underlying characteristics of the people who made that choice and the conditions they faced. When the people who select into a given option differ systematically from the rest of the population, the outcomes one observes are not representative of the option itself, and learning from them requires separating what is attributable to the choice from what is attributable to the chooser. How people perform this separation, and whether they do so at all, shapes their subsequent decisions.

This paper studies how mothers in Sub-Saharan Africa learn from day-of-birth deaths observed in their village, and how they adjust their delivery choices in response. Day-of-birth deaths are ambiguous signals, a death may reflect the quality of the place where the birth occurred, but also the underlying risk of the pregnancy itself. Higher-risk pregnancies are more likely to end in facility births, if complications during labor prompt facility transfer or if mothers with anticipated complications seek facility care. At the same time, facility delivery is more accessible to mothers of higher socioeconomic status, who can more easily sustain the cost of access and may perceive a greater benefit from facility care, and these mothers tend to have ex ante lower-risk pregnancies. As a result, the pool of facility births is selected on observable and unobservable dimensions of risk, and any death observed in a facility combines the contribution of facility quality and the contribution of pregnancy risk. How mothers weigh these two sources when they update their beliefs determines how they adjust their own behavior.

I use data from the Demographic and Health Surveys (DHS), covering the most recent survey round in 38 Sub-Saharan African countries. The final sample includes 272,636 births across more than 20,000 village clusters and spans survey rounds conducted between 2006 and 2024. I exploit within-village variation in the timing, location, and observable risk of preceding births to estimate how mothers respond to locally observed deaths, controlling for village fixed effects and time fixed effects.

The analysis relies on two complementary identification strategies. The first relates changes in the share of facility deliveries in a village to changes in the day-of-birth mortality rate among facility and home births at increasing temporal distance from the index period. This specification provides estimates of the average behavioral response to locally observed mortality and traces its persistence over time. The second is a sequential birth specification that compares consecutive births within a village, relating the delivery location of a birth to the outcome of the immediately preceding birth. This approach exploits very local variation and is particularly well suited to test whether the response to a death varies systematically with the observable risk of the pregnancy that experienced it.

I find that, first, mothers respond strongly to day-of-birth deaths that occur in home births. An increase in the day-of-birth mortality rate among home births in a village is associated with an increase in the share of facility delivery in the following year: the coefficient at the one-year lag is 0.203 ($p < 0.05$), attenuating to 0.078 at two years and no longer distinguishable from zero at three years. Placebo coefficients at the one-year lead are close to zero for both home- and facility-birth mortality, supporting the identifying assumption that the response is not driven by slow-moving trends in village-level facility utilization. The response does not depend on the observable risk of the pregnancy that died at home: mothers update against home delivery whether the pregnancy was observably high or low risk. A death at home is plausibly interpreted as reflecting the absence of medical care, and the response is uniform.

Second, mothers' response to day-of-birth deaths in facility births depends on the observable risk of the pregnancy that experienced the death. The average response to a facility-birth death is small and indistinguishable from zero, but this average masks a sharp gradient. When the deceased baby was born from a pregnancy with no observable risk factors, mothers in the village shift away from facility delivery for subsequent births. When the deceased baby was born from a high-risk pregnancy — characterized by factors such as high parity, young or advanced maternal age, twin birth, no formal education, a short preceding birth interval, or prior miscarriage — mothers shift toward facility delivery. The gradient is monotonic and large: the marginal effect of a facility-birth death on the next facility delivery decision rises from -0.024 when the deceased pregnancy had no observable risk factors to 0.137 when it had six factors. These patterns are consistent with mothers attributing the death to facility quality or to pregnancy risk depending on the observable circumstances. When the deceased pregnancy was observably low-risk, the risk profile does not plausibly explain the death, and the natural inference is that the facility was inadequate; mothers shift away from facility delivery. When the deceased pregnancy was observably high-risk, the risk profile largely explains the death, and the inference about facility quality is weakened; what the death reveals instead is that pregnancies of the kind that produce day-of-birth mortality are present in the local community, raising the perceived need for facility care among mothers who anticipate similar complications.

Third, the increase in facility delivery following a death of a high-risk pregnancy is explained by mothers shifting specifically toward higher-level facilities. I classify delivery locations into hospitals, which include central, provincial, and municipal public hospitals as well as private and NGO hospitals, and lower-level facilities, which include health centers, posts, maternities, private clinics, NGO clinics, and similar structures. Hospitals are typically equipped to perform surgical and emergency obstetric procedures, while lower-level facilities offer routine delivery care but lack the capacity to handle severe complications. When an observably low-risk pregnancy dies in a facility, mothers shift specifically away from hospitals, with no detectable response on the lower-level facility margin. When an observably high-risk pregnancy dies in a facility, mothers shift specifically toward hospitals, again with no movement on the lower-level facility margin. The behavioral response is therefore not a generic reallocation toward facility care, but a targeted reallocation along the hospital margin: mothers move away from hospitals when a death there is read as a signal of inadequate quality, and toward hospitals when a death there is read as a signal that complications requiring high-level care are present in the local community. Taken together, these patterns suggest that mothers attribute locally observed deaths to different sources – facility quality, pregnancy risk, or the level of risk present in the local community – depending on the observable circumstances of the death, and adjust their delivery choices accordingly.

The paper relates to three strands of literature. First, it builds on the literature on peer effects and social learning. Manski (1993) introduced the fundamental identification challenge in separating endogenous, contextual, and correlated effects in social interactions. A subsequent literature has documented social learning from peers' actions (Bandiera and Rasul, 2006; Bursztyn et al., 2014). My paper relates in particular to the literature on learning from observed outcomes, with the agricultural learning literature being particularly influential: Foster and Rosenzweig (1995) model and document how farmers in rural India update beliefs about new technologies by observing the yields of their neighbors, and Conley and Udry (2010) show that farmers in Ghana adjust their input choices toward those of network neighbors whose outcomes were surprisingly successful. This paper contributes to this literature by studying outcome-based social learning in a setting where the signal is intrinsically ambiguous between attributes of the choice and attributes of the chooser. The risk profile of the peer is locally observable, which allows me to test attribution directly: I show that the response to a locally observed death depends systematically on the observable risk characteristics of the pregnancy that experienced it.

Second, the paper relates to a literature on how agents learn about healthcare quality from observed outcomes in the US health sector. Several papers show how observed patient outcomes shape patients' use of healthcare and provider behavior, including referral decisions and delivery mode choices (Chandra et al., 2016; Sarsons, 2017; Singh, 2021; Haye, 2024). Dranove et al. (2003) study selection from the provider side, showing that public disclosure of hospital mortality rates for cardiac surgery led providers to select healthier patients in order to improve their reported outcomes. I document a similar pattern of patient-side updating in a low-income setting, where the inference problem is structurally harder: there are no published performance reports, and the variance in both facility quality and patient risk is much greater. Mothers therefore must rely on locally observed mortality to form beliefs about quality, and they do so by accounting for the observable risk of who experienced the outcome.

Third, the paper contributes to a literature on selection neglect in learning from observed outcomes (Esponda, 2008; Esponda and Vespa, 2018; Barron et al., 2024), which has documented in experimental settings that agents often fail to account for the endogeneity of the samples they observe. The patterns I document offer field evidence in a high-stakes decision: the behavioral response is not naive in the sense of treating every facility-birth death as a quality signal, but depends systematically on whether the pregnancy that died had observable risk factors that could rationalize the death independently of facility quality. Mothers' response is consistent with separating, at least partially, the contribution of facility quality from the contribution of pregnancy risk when interpreting observed deaths.

The remainder of the paper is organized as follows. Section 3.2 presents motivating evidence. Section 3.3 describes the data and the empirical strategy. Section 3.4 presents the results and interprets them through a simple conceptual framework. Section 3.5 concludes.

3.2 Background and motivating evidence

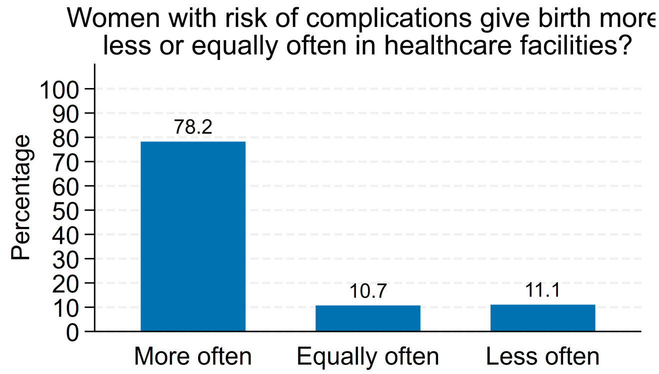
Observed outcomes are the product of two distinct forces: the causal effect of the choice that produced them, and the selection of the people who made that choice. People are not always aware that both forces shape what they observe, and even when they are, disentangling the relative contribution of each to any specific outcome is difficult.

Sub-Saharan Africa has the highest neonatal mortality rate in the world, at 27 deaths per 1,000 live births, and about one third of neonatal deaths occur on the day of birth (WHO, 2024). A day-of-birth death may reflect the quality of the place where the birth occurred, but also the underlying risk of the pregnancy. Higher-risk pregnancies are more likely to end in facility births, if complications during labor prompt facility transfer or if mothers with anticipated complications seek facility care. At the same time, facility delivery is more accessible to mothers of higher socioeconomic status, who can more easily sustain the cost of access and may perceive a greater benefit from facility care, and these mothers tend to have ex ante lower-risk pregnancies. The pool of facility births is therefore selected on observable and unobservable dimensions of risk, and any death observed in a facility combines the contribution of facility quality and the contribution of pregnancy risk.

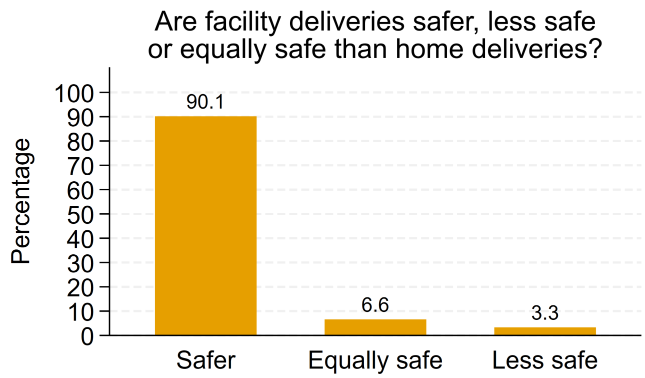
To understand whether mothers themselves are aware of these two forces, I draw on primary survey data collected in 2023 in Zambezia province, Mozambique. The data come from the baseline of the Okokelamo project, a Resilience Food Security Activity funded by USAID and implemented by Save the Children, with the impact evaluation carried out by Innovations for Poverty Action (IPA). The survey instrument was administered to women of reproductive age (15–49) across 26 communities in Zambézia Province and includes a set of questions on perceived safety of facility delivery, on the composition of mothers who deliver in facilities, and on the relationship between place of birth and observed outcomes.

Figure 3.1 reports the distribution of responses to three of these questions. Panel (a) shows that mothers recognize selection into facility delivery: 78.0 percent of respondents report that women with risk of labor complications go more often to give birth in healthcare facilities. Panel (b) shows that mothers also perceive facility delivery as safer: 90.0 percent report that giving birth in a facility is safer than giving birth at home, while only 3.3 percent perceive it as less safe. Together, these two panels indicate that a majority of mothers hold beliefs consistent with both mechanisms: facility delivery is safer in expectation, and pregnancies in facility are riskier in composition.

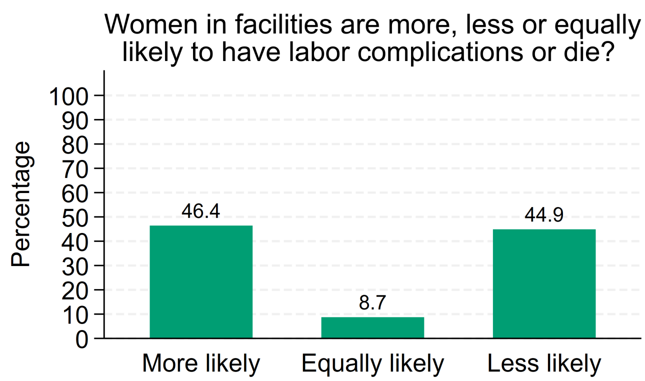
Panel (c) shows the distribution of responses to a question on the observed correlation between place of birth and outcomes. Specifically, the question asks whether women who deliver in facilities experience more or fewer complications and deaths than women who deliver at home. The distribution is mixed: 46.2 percent of respondents report more complications in facilities, 44.9 percent report fewer, and 8.9 percent report the same. The fraction of mothers who recognize the selection mechanism (78.0 percent) is much larger than the fraction who recognize that observed mortality is higher among facility births (46.2 percent). Many mothers



(a) Selection awareness



(b) Perceived safety



(c) Perceived correlation

Figure 3.1. Distribution of responses to three survey questions on perceptions of facility delivery among mothers in 26 communities in Mozambique. Panel (a) reports responses to the question "Women with risk of labor complications go more, less or equally often to give birth in healthcare facilities?", where a response of "more often" indicates awareness that higher-risk pregnancies select into facility delivery. Panel (b) reports responses to the question "Giving birth in a healthcare facility is safer, less safe or equally safe than giving birth at home?". Panel (c) reports responses to the question "Women that give birth in healthcare facilities are more, less or equally likely to have labor complications or die than women that give birth at home?".

Table 3.1. OLS regressions among indicators of selection awareness, perceived safety, and perceived correlation, using survey data from 26 communities in Mozambique. *Selection awareness* is an indicator equal to one if the respondent reports “more often” to the question “Women with risk of labor complications go more, less or equally often to give birth in healthcare facilities?”. *Perceived safety* is an indicator equal to one if the respondent reports “safer” to the question “Giving birth in a healthcare facility is safer, less safe or equally safe than giving birth at home?”. *Perceived correlation* is an indicator equal to one if the respondent reports “more likely” to the question “Women that give birth in healthcare facilities are more, less or equally likely to have labor complications or die than women that give birth at home?”. The sample is restricted to mothers who answered all three questions. Robust standard errors in parentheses.

	(1) Selection awareness	(2) Perceived safety	(3) Perceived correlation
Perceived safety	0.402***		0.0347
	(0.0672)		(0.0702)
Perceived correlation	0.134***	0.0116	
	(0.0321)	(0.0236)	
Selection awareness		0.221***	0.221***
		(0.0425)	(0.0506)
N	568	568	568
Mean of Y	0.79	0.90	0.47

believe in selection and in the safety of facility delivery at the same time, but report that facility births have lower observed mortality than home births. The two mechanisms separately, taken in isolation, are intuitive to a majority of respondents; combining them into a prediction about the observed correlation between place of birth and outcomes is not.

Table 3.1 reports OLS regressions that relate the three indicators to one another. Mothers who are aware of selection are more likely to perceive facility delivery as safer (0.16, $p < 0.01$, Column 2), but selection awareness is not systematically associated with the belief that facility deliveries have more observed complications. Mothers who perceive facility delivery as safer are less likely to believe that facility deliveries have more observed complications (Column 3). The correlations across questions are modest, and they suggest that mothers do not hold a fully coherent integrated view of how selection and safety jointly determine the observed mortality gap.

A more direct test of how mothers process the joint informational content of selection and observed outcomes is reported in Table 3.2. The dependent variable is the indicator for perceiving facility delivery as safer than home delivery. Column (1) shows that selection-aware mothers are 23.5 percentage points more likely to perceive facility delivery as safer than mothers who are not aware of selection ($p < 0.01$). Column (2) interacts selection awareness with reported exposure to delivery complications among relatives and friends, separately for facility and home births. Among mothers who are not aware of selection, exposure to a complication in a facility is associated with a 21.9 percentage point reduction in the probability of perceiving facility delivery as safer ($p < 0.01$): hearing about a facility complication from one's social network strongly downgrades the perceived safety of facility delivery for these mothers. Among mothers who are aware of selection, the net effect of a facility complication on perceived safety is essentially zero (2.6 percentage points, not statistically significant). Mothers who are aware that higher-risk pregnancies deliver in facilities do not appear to update their beliefs about facility safety in response to facility complications, consistent with attributing part of the complication to the composition of who delivers there rather than to facility quality.

A complementary pattern emerges for mothers exposed to complications in both facilities and home births. The net effect of joint exposure is small and statistically indistinguishable from zero for both selection-aware and selection-unaware mothers. When the informational environment is balanced, providing signals from both delivery locations, mothers do not systematically update their beliefs about the relative safety of facility delivery in either direction. This pattern is consistent with the interpretation that the strong downgrading documented for selection-unaware mothers exposed to facility complications alone reflects the salience of one-sided information rather than a uniform response to any negative signal. When mothers observe complications across both locations, the comparative basis for inferring relative safety is restored, and the asymmetry between the two groups disappears.

The patterns documented so far suggest that mothers are aware that selection and facility quality both shape what they observe in their communities, and that integrating these two

Table 3.2. OLS estimates of the association between selection awareness and perceived safety of facility delivery, using primary survey data from 26 communities in Mozambique. The dependent variable is an indicator equal to one if the respondent reports "safer" to the question "Giving birth in a healthcare facility is safer, less safe or equally safe than giving birth at home?". *Aware of selection* is an indicator equal to one if the respondent reports "more often" to the question "Women with risk of labor complications go more, less or equally often to give birth in healthcare facilities?". *Complication in facility* is an indicator equal to one if the respondent answers "yes" to the question "Has any of your relatives and friends had a complicated delivery in a healthcare facility?". *Complication at home* is an indicator equal to one if the respondent answers "yes" to the question "Has any of your relatives and friends had a complicated delivery at home?". The net effect rows report the implied effect of a complication in facility on perceived safety, separately for respondents who are and are not aware of selection, and the effect of exposure to complications in both locations. Standard errors in parentheses.

	Dep. var.: *Facility is safer (0/1)*	
	(1)	(2)
Aware of selection	0.235***	0.219***
	(0.0290)	(0.0328)
Complication in facility		-0.334***
		(0.129)
Aware × Complication in facility		0.360***
		(0.136)
Complication at home		-0.0977
		(0.0896)
Aware × Complication at home		0.0895
		(0.102)
Complication in facility × Complication at home		0.498***
		(0.178)
Aware × Complication in facility × Complication at home		-0.564***
		(0.196)
N	571	571
Effect of facility complication, aware		0.026
Effect of both complications, not aware		0.066
Effect of both complications, aware		-0.049
Mean of Y	0.90	0.90

mechanisms into a single inference about observed outcomes is not straightforward. Mothers who are aware of selection appear to discount the informational content of facility complications, consistent with attributing part of the observed outcome to the composition of who delivers in facilities rather than to facility quality. The Mozambique data offer direct evidence on beliefs that observational data cannot provide, but they are limited in geographic scope and in the granularity of the questions. They cannot speak to how mothers translate these beliefs into delivery choices, nor whether the patterns observed in this specific Mozambican setting generalize to the broader Sub-Saharan context.

To examine the behavioral consequences of these patterns at scale, the rest of the paper turns to the Demographic and Health Surveys (DHS), which are nationally representative household surveys that collect detailed information on fertility, child health, and maternal outcomes. I use the most recent available survey round from 38 Sub-Saharan African countries,[1] which together span survey rounds conducted between 2006 and 2024. Since the DHS records place of delivery only for births occurring in the five years preceding the survey, the sample with information on delivery location includes 317,939 births across 21,079 village clusters. Facility delivery accounts for 70.0 percent of births in the sample, and day-of-birth mortality is rare: at least one such death is observed in 10.4 percent of villages over the observation period, with an average of 0.13 deaths per village.

Figure 3.2 documents a key motivating fact in the DHS data: day-of-birth mortality rates differ between facility and home births across countries, and the gap is highly heterogeneous. In some countries day-of-birth mortality is higher among home-born babies, while in others it is lower. The direction and magnitude of the gap reflect the relative contribution of facility quality and selection in each setting, and the same observed gap is consistent with different combinations of the two forces.

[1]The countries included are: Angola, Benin, Burkina Faso, Burundi, Cameroon, Central African Republic, Chad, Comoros, Democratic Republic of Congo, Republic of Congo, Côte d'Ivoire, Eswatini, Ethiopia, Gabon, Gambia, Ghana, Guinea, Kenya, Lesotho, Liberia, Madagascar, Malawi, Mali, Mauritania, Mozambique, Namibia, Niger, Nigeria, Rwanda, São Tomé and Príncipe, Senegal, Sierra Leone, South Africa, Sudan, Tanzania, Togo, Uganda, Zambia, and Zimbabwe.

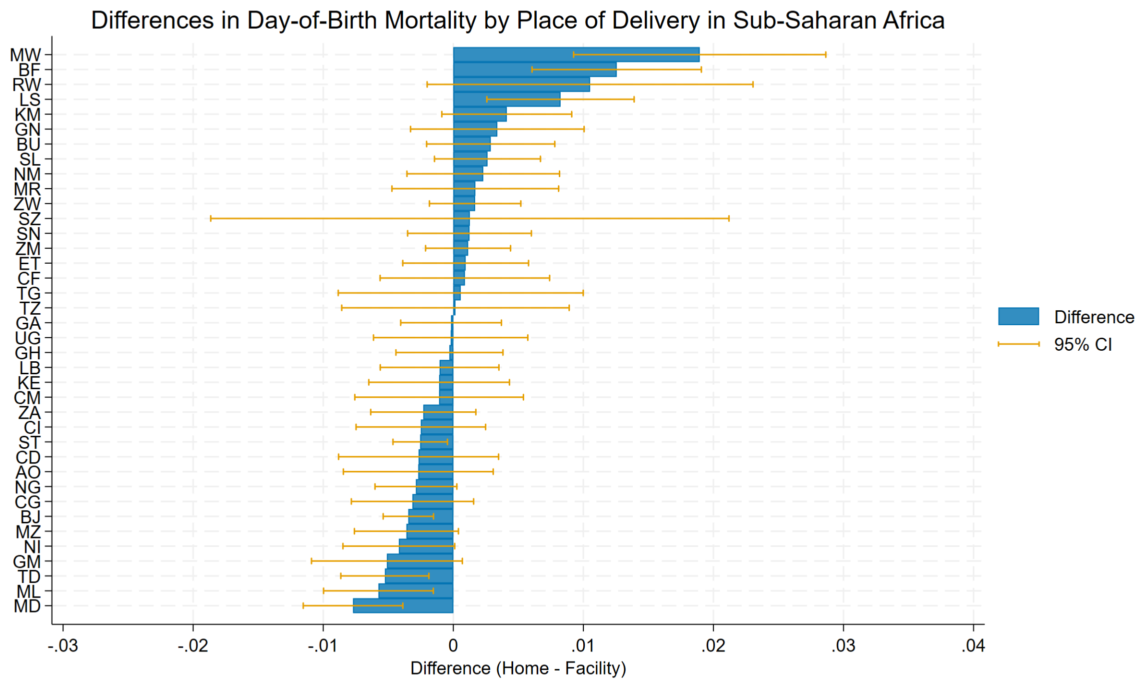


Figure 3.2. Difference in day-of-birth mortality rates between home-born and facility-born babies for each country in the sample, with 95 percent confidence intervals. Positive values indicate higher mortality among home-born babies; negative values indicate higher mortality among facility-born babies.

In the rest of the paper, I exploit within-village variation in the timing, location, and observable risk of preceding births in the DHS to estimate how mothers respond to day-of-birth deaths observed in their communities.

3.3 Empirical strategy

The empirical strategy identifies how mothers respond to outcomes observed in their village, holding constant time-invariant village characteristics. Village fixed effects absorb persistent differences in facility quality, local preferences, and other village-level factors that may correlate with both observed deaths and delivery choices. Time-varying correlated effects cannot be fully ruled out, and I discuss the relevant cases as they arise. I describe the data in Section 3.3.1 and present two complementary specifications: a lagged exposure design (Section 3.3.2) and a sequential birth design (Section 3.3.3).

3.3.1 Data

The DHS records place of delivery for births occurring within five years prior to the survey date, which defines the observation window available for analysis. I classify births as *home births* when delivery occurred at the respondent's home or at another person's home, and as *facility births* when delivery occurred at any type of healthcare facility. Within facility births, I distinguish between two types based on a simplification of the DHS facility classification. I refer to *hospitals* as central, provincial, municipal, private, and NGO hospitals, which typically provide surgical and emergency obstetric capacity. *Lower-level facilities* include health centers, posts, maternities, private clinics, NGO clinics, and other non-hospital structures, which offer routine delivery care but lack capacity to manage severe complications. Hospitals account for 52.2 percent of births in the sample. Summary statistics for the analysis sample are reported in Table A.3.1 in the Appendix.

To capture the observable risk profile of each pregnancy, I construct a risk score summing six indicators of mother-side risk factors: high parity (birth order≥ 5), twin birth, maternal age below 20 or 40 and above, no formal education, a preceding birth interval shorter than 18 months, and a prior miscarriage. The risk score ranges from 0 to 6 and proxies the risk of the pregnancy as inferred from observable characteristics that are plausibly known to other mothers in the same village. I also define *High risk* as an indicator equal to one if the risk score is 2 or higher. Figure A.3.1 in the Appendix shows the distribution of the risk score across births in the analysis sample.

The analytical sample is structured as a panel of DHS clusters, which I refer to throughout as villages, each observed over the five-year DHS window preceding the survey. On average, a village contributes 15 births to the sample. Village-level summary statistics are reported in Table A.3.2 in the Appendix.

3.3.2 Lagged exposure

The first specification relates changes in the share of facility births to changes in the rate of day-of-birth mortality in the village, at varying temporal distance from the index period. Let f_{vt} denote the share of facility births in village v in year t, and let r^F_{vt} and r^H_{vt} denote the day-of-birth mortality rates within facility and home births in the village, respectively, computed as the number of day-of-birth deaths among facility (home) births divided by the number of facility (home) births in the village-year. When no births of a given type occur in the village-year, the corresponding rate is set to zero, reflecting the absence of any informative signal in that category. The specification estimates:

$$\Delta f_{vt} = \sum_{k\in\{1,2,3\}} \left(\beta^F_k\, \Delta r^F_{v,t-k} + \beta^H_k\, \Delta r^H_{v,t-k} + \phi^F_k\, \Delta n^F_{v,t-k} + \phi^H_k\, \Delta n^H_{v,t-k}\right) + \delta_t + \varepsilon_{vt} \qquad (3.1)$$

where Δ denotes the first difference, n^F_{vt} and n^H_{vt} are the numbers of facility and home births in the village in year t, δ_t are year fixed effects, and ε_{vt} is the error term. The regression is weighted by the total number of births in each village-year, and standard errors are clustered at the village level.

The coefficients β_k^F and β_k^H capture how changes in the conditional mortality rate within each delivery location at k-year lags translate into changes in the share of facility births. Expressing the regressors as conditional rates rather than counts isolates the informative content of locally observed outcomes.

This specification exploits the full time series of village outcomes and is robust to truncation of village histories at the beginning of the sample. Controlling separately for $\Delta n^F_{v,t-k}$ and $\Delta n^H_{v,t-k}$ accounts for changes in the number of births of each delivery type, which affect both the sampling variability of the corresponding mortality rate and the mechanical exposure of mothers in the village to observed births of that type. The main limitation is that aggregating over years may dilute the signal of discrete events and reduces the precision with which the timing of behavioral responses can be identified.

3.3.3 Sequential birth

The second specification identifies how delivery choices change from one birth to the next within a village, as a function of the outcome of the immediately preceding birth:

$$f_{vb} = \phi\, f_{v,b-1} + \theta\, d_{v,b-1} + \beta\left(d_{v,b-1} \times f_{v,b-1}\right) + \gamma_v + ym_{vb} + \varepsilon_{vb} \tag{3.2}$$

where b indexes births within a village in chronological order, f_{vb} is an indicator equal to one if the b-th birth in village v occurs in a healthcare facility, $d_{v,b-1}$ is an indicator equal to one if the immediately preceding birth in the village resulted in a day-of-birth death, γ_v are village fixed effects, and ym_{vb} are year-month fixed effects corresponding to the year and month of the b-th birth in village v. The interaction allows the effect of a death to differ depending on whether the preceding birth occurred in a facility.

I extend this specification by interacting all terms with High risk$_{v,b-1}$, an indicator for whether the preceding birth had a risk score of 2 or higher. This tests whether the response to a death depends on the observable risk of the pregnancy that experienced it. To explore whether mothers respond along the quality dimension rather than the binary facility-home choice, I also estimate the same specification with f_{vb} replaced by h_{vb}, an indicator for whether the b-th birth in village v occurred in a hospital.

This approach exploits very local variation, comparing consecutive births within a village, and provides a sharp test of behavioral responses to recent events. The inclusion of the lagged dependent variable allows for persistence in delivery choices and separates changes in behavior from mechanical continuation. The specification is limited to short-run responses and does not capture cumulative effects of exposure to mortality events. It also relies on the assumption that the occurrence of a day-of-birth death in the previous birth is not driven by time-varying unobservables that also affect subsequent delivery choices.

3.4 Results

3.4.1 Average response and persistence

I first estimate equation (3.1) on a panel of village-year observations, where the dependent variable is the change in the share of facility births in the village from one year to the next. The regressors are first differences in the day-of-birth mortality rate within facility and home births, at lags of one through three years, together with a one-year lead used as a placebo. Both rates are computed conditionally on births of the corresponding type, and set to zero in village-years with no births of that type.

Increases in the day-of-birth mortality rate among home births are associated with a large increase in the share of facility deliveries one year later: the coefficient at the one-year lag is 0.203 ($p < 0.05$). To interpret this magnitude, note that a single home-birth day-of-birth death in a village-year corresponds to a home mortality rate between 50 and 100 percent, given the typical one to two home births per village-year; this implies an effect of approximately 10 to 20 percentage points on the share of facility delivery the following year. The effect attenuates to 0.078 at two years and is no longer statistically distinguishable from zero at three years.

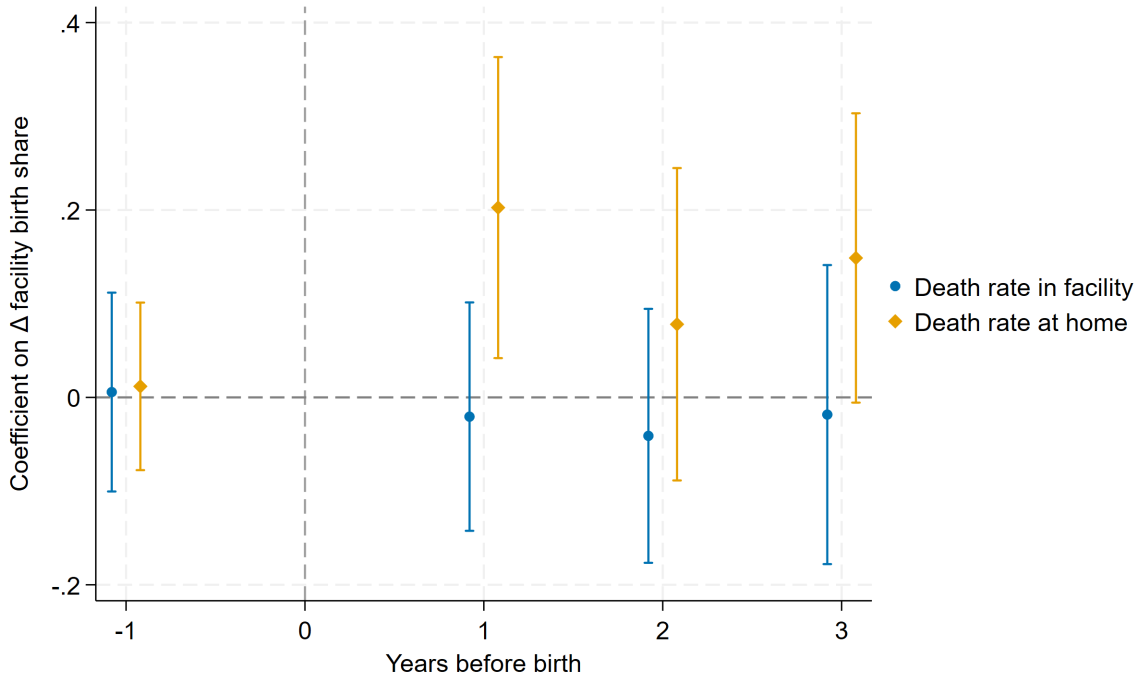


Figure 3.3. Coefficients from the lagged exposure specification in equation (3.1), plotted against the temporal distance of the mortality signal from the index period. Blue circles correspond to changes in the day-of-birth mortality rate among facility births; orange diamonds correspond to changes in the day-of-birth mortality rate among home births. The unit of observation is a village-year; the dependent variable is the first difference in the share of facility births. The coefficient at -1 corresponds to a one-year lead and is used as a placebo test. Point estimates with 95% confidence intervals based on standard errors clustered at the village level.

Increases in the day-of-birth mortality rate among facility births, by contrast, are associated with point estimates close to zero at all lags (-0.021 at one year, -0.041 at two years), none statistically significant at conventional levels.

The placebo coefficients at the one-year lead are 0.012 for home-birth mortality and 0.006 for facility-birth mortality, neither statistically significant. The absence of detectable lead effects supports the identifying assumption that the response is not driven by slow-moving trends in village-level facility utilization correlated with mortality rates.

The lagged exposure specification therefore documents a robust positive response of facility delivery to increases in home-birth mortality and no detectable average response to increases in facility-birth mortality. I turn to the sequential birth specification in the next section to characterize whether the average null response to facility-birth mortality masks heterogeneity by the observable risk of the pregnancy that died.

3.4.2 Heterogeneity by risk of the preceding pregnancy

The sequential birth specification in equation (3.2) compares delivery choices across consecutive births within the same village. The unit of observation is a village-day-of-birth cell for the 29 countries in the sample for which the day of birth is recorded in the DHS, and a village-month cell for the remaining 9 countries[2]. The lagged dependent variable and the lagged death indicator refer to the most recent previous observation within the same village.

To capture the observable risk of each pregnancy, I construct a risk score summing six indicators of mother-side risk factors: high parity (birth order≥ 5), twin birth, maternal age below 20 or 40 and above, no formal education, a preceding birth interval shorter than 18 months, and a prior miscarriage. The risk score ranges from 0 to 6. *High risk* is an indicator equal to one if the risk score is 2 or higher. The indicators rely on information that is plausibly observable to other mothers in the same village, such as the age of the mother, the number and timing of previous births, and visible characteristics of the pregnancy.

Table 3.3 reports estimates. Column (1) reports the baseline specification. The coefficient on Death_{t-1} is 0.035 ($p < 0.05$): a death following a home birth increases the probability of facility delivery for the next observation in the village by 3.5 percentage points. The coefficient on the interaction $\text{Death}_{t-1} \times \text{Facility}_{t-1}$ is -0.025 and not significant. The implied effect of a death following a facility birth, 0.010, is small and indistinguishable from zero. The pattern mirrors the lagged exposure results: a robust positive response to home-birth deaths, no detectable average response to facility-birth deaths.

[2]The day of birth is recorded in the most recent DHS rounds in: Angola, Benin, Burkina Faso, Burundi, Cameroon, Côte d'Ivoire, Democratic Republic of Congo, Ethiopia, Gabon, Gambia, Ghana, Guinea, Kenya, Lesotho, Liberia, Madagascar, Malawi, Mali, Mauritania, Mozambique, Nigeria, Rwanda, Senegal, Sierra Leone, South Africa, Tanzania, Uganda, Zambia, and Zimbabwe. In the remaining nine countries (Central African Republic, Chad, Comoros, Republic of Congo, Eswatini, Namibia, Niger, São Tomé and Príncipe, and Togo), only the year and month of birth are recorded, and observations are aggregated at the village-month level. Results are similar when the analysis is restricted to countries with day-of-birth information.

Table 3.3. Sequential birth estimates of facility delivery following day-of-birth deaths. The dependent variable is Facility_t, an indicator equal to one if the birth at time t occurs in a healthcare facility. Facility_{t-1} and Death_{t-1} are the lagged facility indicator and day-of-birth death indicator for the previous observed birth in the same village. High risk_{t-1} is an indicator equal to one if the previous birth has a risk score of 2 or higher, where the risk score sums six indicators of observable mother-side risk factors: high parity (birth order ≥ 5), twin birth, maternal age below 20 or 40 and above, no formal education, a preceding birth interval shorter than 18 months, and a prior miscarriage. The marginal effect rows in Column (2) report the implied effect of a day-of-birth death following a facility birth, evaluated at $\text{High risk}_{t-1} = 0$ and $\text{High risk}_{t-1} = 1$. All specifications include village and year-month fixed effects. Standard errors are clustered at the village level.

	Dep. var.: *$Facility_t$ (0/1)*	
	(1)	(2)
$\text{Facility}_{\text{t-1}}$	-0.0625***	-0.0652***
	(0.00250)	(0.00276)
$\text{Death}_{\text{t-1}}$	0.0353**	0.0552**
	(0.0172)	(0.0265)
$\text{Death}_{\text{t-1}} \times \text{Facility}_{\text{t-1}}$	-0.0246	-0.0648**
	(0.0193)	(0.0288)
$\text{Death}_{\text{t-1}} \times \text{High risk}_{\text{t-1}}$		-0.0374
		(0.0350)
$\text{Facility}_{\text{t-1}} \times \text{High risk}_{\text{t-1}}$		0.00775**
		(0.00353)
$\text{Death}_{\text{t-1}} \times \text{Facility}_{\text{t-1}} \times \text{High risk}_{\text{t-1}}$		0.0926**
		(0.0397)
ME of facility death, low risk		-0.010
ME of facility death, high risk		0.046***
N	272,636	272,636
N of clusters	20,174	20,174
Mean of Y	.71	.71
Year-month fixed effect	Yes	Yes
Village fixed effect	Yes	Yes

Column (2) interacts all terms with High risk_{t-1}. The triple interaction $\text{Death}_{t-1} \times \text{Facility}_{t-1} \times \text{High risk}_{t-1}$ is 0.092 ($p < 0.05$). Evaluated at $\text{High risk}_{t-1} = 0$, the marginal effect of a facility-birth death is -0.010 and not significant. Evaluated at $\text{High risk}_{t-1} = 1$, the marginal effect is 0.046 ($p < 0.01$). Mothers do not detectably change behavior when an observably low-risk pregnancy dies in a facility, but shift toward facility delivery when an observably high-risk pregnancy dies in a facility.

Figure 3.4 disaggregates this gradient using the continuous risk score. The figure plots the marginal effect of a home-birth death (left panel) and a facility-birth death (right panel) across the range of the risk score from 0 to 6, computed from a specification that interacts equation (3.2) with the continuous score.

The gradient on facility-birth deaths is monotonic, with the marginal effect rising from -0.024 at risk = 0 to 0.137 at risk = 6. The effect is not statistically distinguishable from zero at the bottom of the risk distribution and becomes positive and significant for higher values of the risk score. The marginal effect of a home-birth death, by contrast, is essentially flat across the risk distribution at around 0.037. At the upper end of the risk distribution, the response to a facility-birth death converges to the response to a home-birth death.

Two features of the pattern stand out. First, the response to a home-birth death is uniform across the risk distribution: mothers shift toward facility delivery after a home-birth death regardless of whether the deceased pregnancy was observably low- or high-risk. Second, the response to a facility-birth death depends sharply on the observable risk of the pregnancy. A death of a low-risk pregnancy in a facility is followed by a shift away from facility delivery; a death of a high-risk pregnancy in a facility is followed by a shift toward facility delivery, with effects that grow with the observable risk.

These patterns are consistent with mothers attributing locally observed deaths to different sources depending on the observable circumstances. A death at home is interpreted as reflecting the absence of medical care; the response is uniform because home delivery is interpreted as unsafe regardless of the risk profile of the pregnancy that died there. A death of a low-risk pregnancy in a facility is harder to attribute to the underlying risk of the pregnancy and is more naturally interpreted as a signal that facility care is inadequate. A death of a high-risk pregnancy in a facility is plausibly attributable to the difficulty of the case and reveals that complications

Table 3.4. Sequential birth estimates of delivery location following day-of-birth deaths, with lower-level facilities and hospitals treated as separate outcomes. The dependent variable in Columns (1) and (2) is Lower level facility$_t$, an indicator equal to one if the birth at time t occurs in a health center, health post, maternity, private clinic, private medical center, NGO clinic, or other non-hospital facility. The dependent variable in Columns (3) and (4) is Hospital$_t$, an indicator equal to one if the birth at time t occurs in a central, provincial, or municipal public hospital, a private hospital, or an NGO hospital. Facility$_{t-1}$ and Death$_{t-1}$ are the lagged facility indicator and day-of-birth death indicator for the previous observed birth in the same village. High risk$_{t-1}$ is an indicator equal to one if the previous birth has a risk score of 2 or higher. The marginal effect rows in Columns (2) and (4) report the implied effect of a day-of-birth death following a facility birth on the respective outcome, evaluated at High risk$_{t-1} = 0$ and High risk$_{t-1} = 1$. All specifications include village and year-month fixed effects. Standard errors are clustered at the village level.

	Lower level facility $_t$ (0/1)		*Hospital$_t$ (0/1)*	
	(1)	(2)	(3)	(4)
Facility$_{t-1}$	-0.0162***	-0.0166***	-0.0464***	-0.0487***
	(0.00171)	(0.00189)	(0.00233)	(0.00261)
Death$_{t-1}$	0.00282	0.0204	0.0325**	0.0348
	(0.0120)	(0.0201)	(0.0158)	(0.0246)
Death$_{t-1}$ × Facility$_{t-1}$	0.00945	-0.0104	-0.0340*	-0.0542**
	(0.0141)	(0.0220)	(0.0186)	(0.0275)
Death$_{t-1}$ × High risk$_{t-1}$		-0.0334		-0.00382
		(0.0242)		(0.0318)
Facility$_{t-1}$ × High risk$_{t-1}$		0.00118		0.00697**
		(0.00251)		(0.00326)
Death$_{t-1}$ × Facility$_{t-1}$ × High risk$_{t-1}$		0.0396		0.0528
		(0.0282)		(0.0374)
N	272,636	272,636	272,636	272,636
ME of facility death, low risk		0.010		-0.019*
ME of facility death, high risk		0.016		0.030*
N of clusters	20,174	20,174	20,174	20,174
Mean of Y	.19	.19	.52	.52
Year-month fixed effect	Yes	Yes	Yes	Yes
Village fixed effect	Yes	Yes	Yes	Yes

carrying high mortality risk occur in the local community.

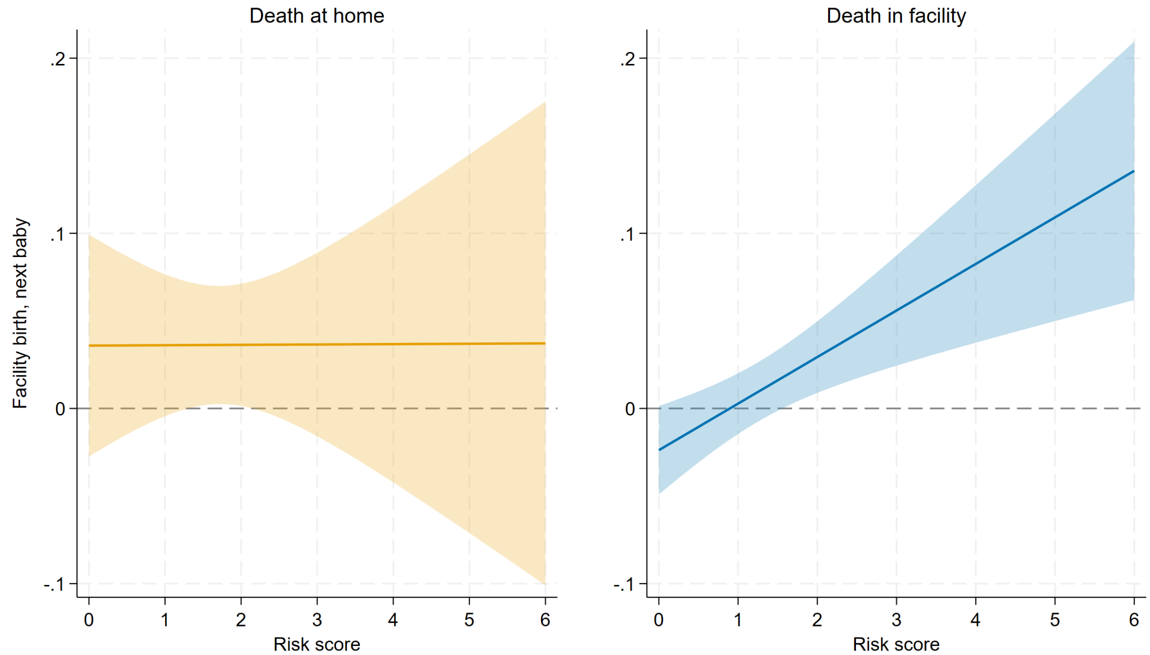


Figure 3.4. Marginal effect of a day-of-birth death on subsequent facility delivery, plotted across the range of the risk score from 0 to 6. The left panel shows the marginal effect of a death following a home birth; the right panel shows the marginal effect of a death following a facility birth. The risk score sums six indicators of mother-side risk factors of the previous birth. Marginal effects are computed from a continuous version of the specification in Column (2) of Table 3.3, in which the risk score enters linearly rather than as a binary indicator. Shaded regions denote 95% confidence intervals based on standard errors clustered at the village level.

3.4.3 Shift toward higher-level facilities

The previous subsection treats facility delivery as a binary outcome, but the increase in facility delivery following a death of a high-risk pregnancy admits more than one interpretation. Mothers may be moving toward any facility indifferently, or they may be moving specifically toward facilities with greater capacity to manage complications. To distinguish these two possibilities, I classify the delivery locations in the DHS into two groups. The first group, which I refer to as *hospitals*, includes central, provincial, and municipal public hospitals, as well as private and NGO hospitals. The second group, which I refer to as *lower-level facilities*, includes health centers, health posts, maternities, private clinics and medical centers, NGO clinics, and other non-hospital facilities. Hospitals are typically equipped to perform surgical and emergency obstetric procedures, while lower-level facilities offer routine delivery care but lack the capacity to handle severe complications.

Table 3.4 reports estimates of the sequential birth specification with hospital delivery and lower-level facility delivery as separate outcomes. Columns (1) and (2) report the specification for lower-level facility delivery; Columns (3) and (4) report the same specification for hospital delivery. The structure mirrors Table 3.3.

Columns (1) and (2) show that the behavioral response operates weakly, if at all, on the lower-level facility margin. The marginal effect of a facility-birth death on lower-level facility delivery is 0.010 at low risk and 0.016 at high risk, neither statistically distinguishable from zero. Mothers do not appear to redirect toward lower-level facilities in response to facility-birth deaths, regardless of the observable risk of the deceased pregnancy.

Columns (3) and (4) show that the response is concentrated on the hospital margin. The marginal effect of a facility-birth death on hospital delivery is -0.019 at low risk ($p < 0.10$): when an observably low-risk pregnancy dies in a facility, mothers shift specifically away from hospitals. The marginal effect at high risk is 0.030 ($p < 0.10$): when an observably high-risk pregnancy dies in a facility, mothers shift specifically toward hospitals.

The contrast across the two outcomes is informative. The shift in delivery behavior following facility-birth deaths is not a reallocation toward facility care in general, but a reallocation specifically along the hospital margin. When mothers move away from facility care after a low-risk pregnancy death in a facility, they move away from hospitals, not toward lower-level facilities. When they move toward facility care after a high-risk pregnancy death, they move toward hospitals, not toward lower-level facilities. The increase in facility delivery documented in Table 3.3 for high-risk deaths is therefore better described as an increase in hospital delivery: mothers respond by seeking facilities with the capacity to handle severe complications, not by increasing demand for facility care of any type.

3.5 Conclusion

This paper examines how mothers in Sub-Saharan Africa update their delivery choices in response to day-of-birth deaths observed in their village. Using primary survey data from Mozambique together with the Demographic and Health Surveys covering 38 Sub-Saharan African countries, I document patterns consistent with mothers separating, at least partially, the contribution of facility quality from the contribution of pregnancy risk when interpreting locally observed outcomes.

Three main findings emerge from the empirical analysis. First, mothers respond strongly to day-of-birth deaths that occur in home births, shifting toward facility delivery in a way that does not depend on the observable risk of the deceased pregnancy. The response to home-birth deaths is consistent with attributing the death to the absence of medical care. Second, the response to deaths in facility births exhibits a sharp gradient in the observable risk of the deceased pregnancy. When a low-risk pregnancy dies in a facility, mothers shift away from facility delivery; when a high-risk pregnancy dies in a facility, mothers shift toward it. The transition from a negative to a positive response is monotonic, with marginal effects rising from -0.024 to 0.137 across the range of observable risk factors. Third, the response to facility-birth deaths is concentrated along the hospital margin: mothers move specifically away from hospitals when the deceased pregnancy was low-risk, and specifically toward hospitals when the deceased pregnancy was high-risk. The behavioral response is not a generic reallocation toward or away from facility care, but a targeted shift in the level of care chosen.

These findings have implications for how we understand information frictions in healthcare-seeking decisions. In settings where data on provider performance are not publicly available, mothers form beliefs through local observation, and those observations are shaped by the same selection processes that determine who delivers in which type of facility. The patterns documented here suggest that mothers respond to this informational environment with some sophistication: they do not treat every facility-birth death as a signal of poor facility quality, but condition their inference on the observable risk of the pregnancy that experienced the death. The survey evidence from Mozambique reinforces this interpretation. Mothers who are aware that higher-risk pregnancies select into facility delivery do not downgrade their perception of facility safety in response to facility complications, while mothers without this awareness do.

Several limitations of the analysis are worth highlighting. The empirical strategy relies on within-village variation controlling for village and time fixed effects, and time-varying unobservables that correlate with both the occurrence of deaths and subsequent delivery choices cannot be fully ruled out. The DHS data also do not allow direct observation of beliefs; the conceptual framework I propose is consistent with the behavioral patterns observed, but alternative models of belief formation may generate similar predictions. The Mozambique survey data provide direct evidence on beliefs but are geographically limited and collected at a single point in time, leaving open how these beliefs evolve.

The findings raise two sets of questions that future work could address. The first concerns what the shift toward hospitals after the death of a high-risk pregnancy in a facility actually reflects. This shift could operate through two distinct channels. One is an information channel: the death informs mothers in the village that high-risk pregnancies of the kind that produce day-of-birth mortality are present in their community, and they update toward hospital delivery accordingly. The other is a structural complementarity between facility quality and pregnancy risk, which I document directly in Budillon (forthcoming): for high-risk pregnancies, the survival benefit of hospital delivery (relative to a lower-level facility) is mechanically larger, regardless of any new information. The patterns I document here are consistent with both channels operating together, and the data in this paper do not allow me to separate them. Disentangling the two matters for policy: if the response is primarily informational, interventions that improve mothers' access to information about facility performance would expand appropriate care; if it is primarily structural, the binding constraint is the capacity and geographic availability of hospitals.

The second set of questions concerns how information about locally observed deaths actually circulates in the village. The empirical strategy treats a death as if it were equally observable to every mother in the village, but in practice information may travel through social ties, kin networks, or community gatherings, and some mothers may receive the signal more reliably than others. The structure of these communication channels likely shapes which villages respond more strongly to a given observed death, and could inform the design of interventions that work with, rather than around, the local channels through which health information already spreads.

A.1 Chapter 1 Appendix

A.1.1 Construction of indexes

Baseline mortality index

Facility births provide emergency interventions that are particularly valuable when complications arise, which motivates examining whether the returns to facility birth are largest for those with the greatest underlying need. To capture this, I construct an index of baseline mortality risk by estimating the following logistic regression of neonatal mortality on a set of pre-birth characteristics of the mother and baby:

$$\text{logit}\left(\Pr(\text{Neonatal mortality}_i = 1)\right) = \beta_0 + X_i\beta + \varepsilon_i \tag{3}$$

and computing the predicted probability as

$$\widehat{\text{Neonatal mortality}}_i = \widehat{\beta}_0 + X_i\widehat{\beta}. \tag{4}$$

The vector X_i includes the baby's year and month of birth, gender, the mother's age at birth, years of education, whether she experienced a terminated pregnancy before, and the household's source of drinking water.

Resources index: multiple correspondence analysis

To measure facility resources, I construct an index using Multiple Correspondence Analysis applied to SPA facility characteristics. The SPA items are categorical and record both availability and functioning, for example not present, present but not functioning, functioning as observed by the enumerator, functioning as reported by staff, or do not know. I treat missing values as a separate category. MCA transforms each categorical item into a set of indicator variables, one for each response category, and then extracts common dimensions. The analysis uses all hospitals indicated as closest in my sample, and I take the first dimension as the resources index. I include SPA items on facility cleanliness, medications, services, general equipment, newborn equipment, newborn practices, obstetric guidelines, delivery practices, delivery supplies, Prevention of Mother-To-Child Transmission (PMTCT) supplies, surgical supplies, outpatient supplies, immunization supplies, sick child routine, and sick child practices. The full list of variables for each block is reported in Table A.1.1.

Table A.1.1. Variables included in the facility resources index. Data from the Malawi Service Provision Assessment (SPA) 2013. PMTCT stands for Prevention of Mother-To-Child Transmission of HIV.

Block	**Variables**
Cleanliness	v154-v154h
Medications	v534a-v534o; v903_02-v924_05
Services	vt201-vt1501
General equipment	v533b-v533k
Newborn equipment	v536a-v536f
Newborn practices	v506c-v507m
Obstetric guidelines	v537b-v537e
Delivery practices	v554b-v567
Delivery supplies	v530-v531xo
Delivery PMTCT supplies	v1281a1-v1281c
PMTCT supplies	v1276-v1277w
Surgical supplies	v1702a-v1703e
Outpatient supplies	v165a-v168w
Immunization supplies	v231a1-v231xg
Sick child routine	v247a-v247h
Sick child practices	v264a1-v267d

Resources index: additive index

As a robustness check, I construct a simple additive index using the same item blocks listed in Section A.1.1. The index captures whether facility j meets essential criteria k:

$$\text{ResourcesIndex}_j = \frac{1}{K}\sum_{k=1}^{K} \mathbb{1}(\text{Facility } j \text{ has item } k) \tag{5}$$

where K is the total number of quality-related items or services assessed, and the indicator function $\mathbb{1}$ equals one if the facility possesses a given item or provides the corresponding service.

Each criterion receives equal weight and $\text{ResourcesIndex}_j \in [0,1]$.

A.1.2 Additional tables and figures

Table A.1.2. Correlation between neonatal mortality and distance to healthcare facilities. The unit of analysis is a DHS cluster in Malawi. Distance is measured as the path length to the nearest healthcare facility. The sample includes 850 facilities distributed across 69 grids; 5 grids with singleton observations are excluded.

	Neonatal mortality [0-1]			
	(1)	(2)	(3)	(4)
Distance to any health facility (100 km)	0.00596		0.00647	
	(0.0129)		(0.0125)	
Distance to a facility with a doctor (100 km)		-0.00130***		0.00995**
		(0.000307)		(0.00498)
N	64619	64619	65005	65005
Clusters	414	414	419	419
Mean of Y	0.0241	0.0241	0.0240	0.0240
50km grid FE	No	No	Yes	Yes

Table A.1.3. Baby, mother, and village characteristics by place of delivery, including missing values. Column 1 reports mean values for births occurring at home, column 2 for births in facilities, and column 3 the difference between the two. Column 4 reports mean values for births where the place of delivery is not recorded, and column 5 shows the difference between births with recorded delivery location (home or facility) and those with missing delivery location. Standard deviations are in parentheses.

	(1) Home	(2) Facility	(3) Difference (1) - (2)	(4) Missing facility	(5) Difference (1)&(2) - (4)
Baby					
Birth order	3.88	3.14	0.74***	2.87	0.32***
	(2.29)	(2.05)	(0.07)	(1.88)	(0.02)
Infant mortality (0/1)	0.05	0.04	0.02***	0.06	-0.02***
	(0.23)	(0.19)	(0.01)	(0.24)	(0.00)
Neonatal mortality (0/1)	0.04	0.02	0.02***	0.02	-0.00
	(0.20)	(0.15)	(0.01)	(0.16)	(0.00)
Received antenatal care (0/1)	0.90	0.99	-0.09***	0.95	0.04**
	(0.30)	(0.11)	(0.01)	(0.23)	(0.02)
Received postnatal care (0/1)	0.43	0.45	-0.02	0.51	-0.07*
	(0.49)	(0.50)	(0.02)	(0.50)	(0.04)
Mother					
Mother's age	27.01	25.91	1.10***	23.75	2.23***
	(7.20)	(6.66)	(0.22)	(5.77)	(0.06)
Years of education	4.07	5.80	-1.73***	4.58	1.11***
	(3.18)	(3.58)	(0.10)	(3.67)	(0.03)
Total children ever born	4.24	3.37	0.87***	5.37	-1.94***
	(2.31)	(2.08)	(0.07)	(2.18)	(0.02)
Visited health facility in last 12 months (0/1)	0.70	0.75	-0.04***	0.65	0.09***
	(0.46)	(0.43)	(0.01)	(0.48)	(0.00)
Village					
Rural (0/1)	0.92	0.83	0.09***	0.84	0.00
	(0.26)	(0.37)	(0.01)	(0.37)	(0.00)
Distance to hospital (km)	9.25	6.71	2.54***	6.86	0.01
	(9.79)	(6.28)	(0.30)	(6.69)	(0.06)
Closest facility has a doctor (0/1)	0.15	0.21	-0.06***	0.21	-0.01***
	(0.35)	(0.40)	(0.01)	(0.41)	(0.00)
N	1,096	15,780	16,876	48,129	65,005

Figure A.1.1. Example of distance measurement from a DHS cluster to its nearest health facility offering delivery services. The map is extracted from OpenStreetMap (OSM) and illustrates the shortest travel route along the road network. The green pin marks the location of a village (DHS cluster) and the red pin marks the nearest health facility. The distance is calculated as the length of the route highlighted in purple.

Table A.1.4. First-stage results under different specifications. The dependent variable is a binary indicator equal to 1 if the child was born in a healthcare facility. Time fixed effects are indicators for each month and year pair. Seasonality fixed effects use indicators for each grid and calendar month. Monthly rain controls correspond to the average rainfall in each of the twelve months preceding birth. Expected rain controls refer to the average daily rainfall in a two-week window around the expected delivery date, calculated as the historical average over the five years preceding birth. Standard errors are clustered at the grid-month level.

	Dep. var.: Facility birth (0/1)				
	(1)	(2)	(3)	(4)	(5)
Rain (m) × Distance (5km)	-0.192**	-0.192**	-0.258***	-0.259**	-0.258***
	(0.0828)	(0.0801)	(0.0972)	(0.114)	(0.0897)
Rain (m)	-0.217	-0.0277	0.0643	0.0241	-0.0124
	(0.201)	(0.227)	(0.273)	(0.286)	(0.270)
Distance (5km)	-0.0143***	-0.0144***	-0.0188***	-0.0516***	-0.0550***
	(0.00283)	(0.00283)	(0.00304)	(0.0113)	(0.0121)
N	12151	12151	12107	12106	12106
N of clusters	792	792	748	748	748
Mean of Y	0.928	0.928	0.928	0.928	0.928
F (Kleibergen-Paap)	5.4	5.8	7.1	5.1	8.3
Time FE	No	Yes	Yes	Yes	Yes
Seasonality FE	No	No	Yes	Yes	Yes
Monthly rain (past year) controls	No	No	No	Yes	Yes
Expected rain controls	No	No	No	No	Yes

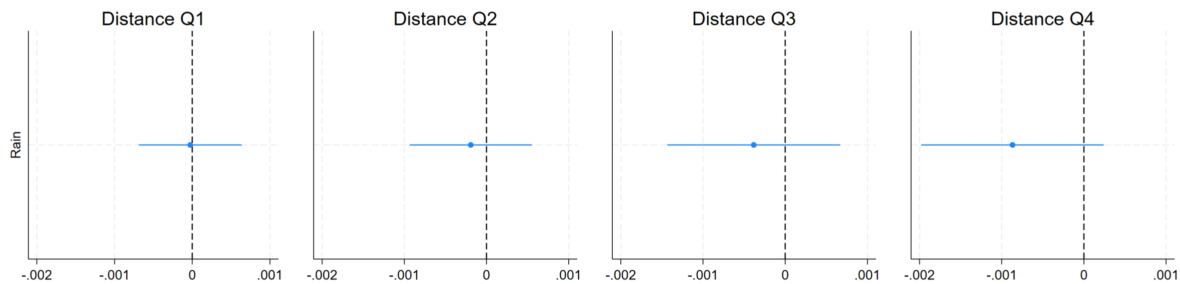


Figure A.1.2. Non-parametric first stage. Coefficients of interaction of rain and indicators for distance quartiles on facility birth. The specification also includes indicators for the main effects of the rain and distance bins.

Table A.1.5. Reduced-form estimates for neonatal mortality. The dependent variable is a binary indicator equal to 1 if the baby died within 28 days of birth. High risk is a dummy equal to 1 if the standardized log-odds from a logistic regression of neonatal mortality on pre-birth covariates is above the median. Doctor equals 1 if the closest facility has at least one doctor on staff. Controls include past and expected rainfall, as well as their interactions with distance. Seasonality and time fixed effects are included. Standard errors are clustered at the grid-month level.

	Dep. var.: Neonatal mortality (0/1)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	First stage *Facility birth (0/1)*						
Rain (m) × Distance (5km)	-0.280***	-0.236**	-0.0147	-0.0578	-0.0288	-0.0704*	-0.0643*
	(0.0877)	(0.0992)	(0.0288)	(0.0351)	(0.0293)	(0.0374)	(0.0363)
Rain (m) × Distance (5km) × High mortality risk = 1		-0.673*		0.133**		0.128**	0.109*
		(0.362)		(0.0591)		(0.0594)	(0.0589)
Rain (m) × Distance (5km) × Doctor = 1		0.826			0.463**	0.452**	0.126
		(0.740)			(0.194)	(0.194)	(0.158)
Rain (m) × Distance (5km) × Doctor = 1 × High mortality risk = 1		-1.339					0.629*
		(2.011)					(0.348)
N	12106	12106	46567	46567	46567	46567	46567
N of clusters	748	748	811	811	811	811	811
Mean of Y	0.928	0.928	0.023	0.023	0.023	0.023	0.023
Effect p25 to p75	-.006 ***	-.005 **	0	-.001	-.001	-.002 *	-.001 *
Effect p25 to p75 × High mortality risk		-.019 ***		.002		.001	.001
Effect p25 to p75 × Doctor		.013			.009 **	.008 **	.001
Effect p25 to p75 × High mortality risk × Doctor		-.03					.017 **

Table A.1.6. Second-stage results for future healthcare visits: postnatal period excluded. The dependent variable is a binary indicator equal to 1 if the mother reported visiting a healthcare facility in the past 12 months. The outcome is measured at the mother level and observations are weighted by the inverse of the number of births per mother. Resources is a composite index of facility characteristics, grouped into terciles. Controls include past and expected rainfall, as well as their interactions with distance. Seasonality and time fixed effects are included. Standard errors are clustered at the mother level.

	All births		Last 5 years births					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Reduced form *Visited health facility last 12 months (0/1)*		**First stage** *Facility birth (0/1)*		**IV** *Visited health facility last 12 months (0/1)*		**Reduced form** *Visited health facility last 12 months (0/1)*	
Rain (m) × Distance (5km)	0.198*	0.389***	-0.396***	-0.525***			0.129	0.388**
	(0.119)	(0.146)	(0.111)	(0.134)			(0.201)	(0.189)
Rain (m) × Distance (5km) × Resources = Intermediate		-0.439		0.480*				-0.906*
		(0.273)		(0.263)				(0.463)
Rain (m) × Distance (5km) × Resources = High		-0.605*		0.583				0.0442
		(0.335)		(0.425)				(0.725)
Facility birth (0/1)					-0.325	-1.651		
					(0.493)	(2.970)		
Facility birth (0/1) × Resources = Intermediate						2.108		
						(46.99)		
Facility birth (0/1) × Resources = High						18.90		
						(59.76)		
N	43399	43399	8974	8974	8974	9420	8974	8974
N of clusters	12045	12045	7731	7731	7731	7983	7731	7731
Mean of Y	0.667	0.667	0.923	0.923	0.730	0.730	0.730	0.730
Effect p25 to p75	.004 *	.009 ***	-.009 ***	-.012 ***			.003	.009 **
Effect p25 to p75 × Resources = Intermediate		-.001		-.001				-.012
Effect p25 to p75 × Resources = High		-.005		.001				.01
F (Kleibergen-Paap)					12.819	0.026		

Table A.1.7. Effect of access to facility birth on fertility. The dependent variable is a binary indicator equal to 1 if the baby has a younger sibling. Column (1) includes all observations. Column (2) restricts the sample to births occurring at least three years before the survey, while column (3) restricts to mothers aged 40 or older, representing two possible ways to look at the effects of the shock conditional on realized fertility. Controls include past and expected rainfall, as well as their interactions with distance. Seasonality and time fixed effects are included. Standard errors are clustered at the grid-month level.

	Dep. var.: has a younger sibling (0/1)		
	(1) All observations	(2) Children at least 3 yo	(3) Mother at least 40 yo
Rain (m) X Distance (5km)	0.0758	-0.404	0.229
	(0.0961)	(0.287)	(0.211)
N	44831	5465	13818
N of clusters	811	659	727
Mean of Y	0.750	0.0884	0.834
F (Kleibergen-Paap)	0.6	2.0	1.2
Effect p25 to p75	0.00167	-0.00950	0.00511

Table A.1.8. Second-stage results for neonatal mortality using clinical-only variables to predict baseline mortality risk. The dependent variable is a binary indicator equal to 1 if the baby died within 28 days of birth. High risk is a dummy equal to 1 if the standardized log-odds from a logistic regression of neonatal mortality on clinical-only pre-birth covariates is above the median. Doctor equals 1 if the closest facility has at least one doctor on staff. Controls include past and expected rainfall, as well as their interactions with distance. Seasonality and time fixed effects are included. Standard errors are clustered at the grid-month level.

	Dep. var.: Neonatal mortality (0/1)						
	(1) First stage *Facility birth (0/1)*	(2)	(3)	(4)	(5)	(6)	(7)
Rain (m) × Distance (5km)	-0.280***	-0.289	-0.0147	-0.0343	-0.0288	-0.0527	-0.0364
	(0.0877)	(0.287)	(0.0288)	(0.0494)	(0.0293)	(0.0503)	(0.0506)
Rain (m) × Distance (5km) × High mortality risk (0/1)		0.0324		0.0276		0.0331	0.0115
		(0.315)		(0.0551)		(0.0553)	(0.0556)
Rain (m) × Distance (5km) × Doctor (0/1)		0.584			0.463**	0.468**	-0.0699
		(0.686)			(0.194)	(0.195)	(0.164)
Rain (m) × Distance (5km) × Doctor (0/1) × High mortality risk (0/1)		-0.817					0.838***
		(1.730)					(0.321)
N	12106	12106	46567	46567	46567	46567	46567
N of clusters	748	748	811	811	811	811	811
Mean of Y	0.928	0.928	0.023	0.023	0.023	0.023	0.023
Effect p25 to p75	-.006 ***	-.006	0	-.001	-.001	-.001	-.001
Effect p25 to p75 × High mortality risk		-.005 ***		0		0	-.001
Effect p25 to p75 × Doctor		.006			.009 **	.009 **	-.002
Effect p25 to p75 × High mortality risk × Doctor		-.01					.016 ***

Table A.1.9. Second-stage results for neonatal mortality using a linear probability model to predict baseline mortality risk. The dependent variable is a binary indicator equal to 1 if the baby died within 28 days of birth. High risk equals 1 if the standardized fitted values from a linear probability model of neonatal mortality on pre-birth covariates exceed the median. Doctor equals 1 if the closest facility has at least one doctor on staff. Controls include past and expected rainfall, as well as their interactions with distance. Seasonality and time fixed effects are included. Standard errors are clustered at the grid-month level

	Dep. var.: Neonatal mortality (0/1)						
	(1) First stage *Facility birth (0/1)*	(2)	(3)	(4)	(5)	(6)	(7)
Rain (m) × Distance (5km)	-0.280***	-0.274***	-0.0147	-0.0314	-0.0288	-0.0442	-0.0423
	(0.0877)	(0.0906)	(0.0288)	(0.0302)	(0.0293)	(0.0312)	(0.0310)
Rain (m) × Distance (5km) × High mortality risk (0/1)		-0.454		0.0601		0.0536	0.0459
		(0.352)		(0.0660)		(0.0661)	(0.0670)
Rain (m) × Distance (5km) × Doctor (0/1)		0.303			0.463**	0.461**	0.320*
		(0.776)			(0.194)	(0.195)	(0.193)
Rain (m) × Distance (5km) × Doctor (0/1) × High mortality risk (0/1)		0.0452					0.267
		(1.950)					(0.348)
N	12106	12106	46567	46567	46567	46567	46567
N of clusters	748	748	811	811	811	811	811
Mean of Y	0.928	0.928	0.023	0.023	0.023	0.023	0.023
Effect p25 to p75	-.006 ***	-.006 ***	0	-.001	-.001	-.001	-.001
Effect p25 to p75 × High mortality risk		-.016 **		.001		0	0
Effect p25 to p75 × Doctor		.001			.009 **	.009 **	.006
Effect p25 to p75 × High mortality risk × Doctor		-.008					.013 *

Table A.1.10. Second-stage results for future healthcare visits using additive resources index. The dependent variable is a binary indicator equal to 1 if the mother reported visiting a healthcare facility in the past 12 months. The outcome is measured at the mother level and observations are weighted by the inverse of the number of births per mother. Resources is a composite index of facility characteristics, grouped into terciles. Controls include past and expected rainfall, as well as their interactions with distance. Seasonality and time fixed effects are included. Standard errors are clustered at the mother level.

	All births		Last 5 years births					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Reduced form *Visited health facility last 12 months (0/1)*		**First stage** *Facility birth (0/1)*		**IV** *Visited health facility last 12 months (0/1)*		**Reduced form** *Visited health facility last 12 months (0/1)*	
Rain (m) × Distance (5km)	0.181	0.413***	-0.372***	-0.503***			0.137	0.471**
	(0.121)	(0.145)	(0.111)	(0.130)			(0.210)	(0.185)
Rain (m) × Distance (5km) × Resources = Intermediate		-0.531*		0.461*				-1.090**
		(0.274)		(0.262)				(0.471)
Rain (m) × Distance (5km) × Resources = High		-0.740**		0.644				-0.390
		(0.337)		(0.417)				(0.723)
Facility birth (0/1)					-0.370	-1.651		
					(0.542)	(2.970)		
Facility birth (0/1) × Resources = Intermediate						2.108		
						(46.99)		
Facility birth (0/1) × Resources = High						18.90		
						(59.76)		
N	43850	43850	9420	9420	9420	9420	9420	9420
N of clusters	12154	12154	7983	7983	7983	7983	7983	7983
Mean of Y	0.668	0.668	0.925	0.925	0.730	0.730	0.730	0.730
Effect p25 to p75	.004	.009 ***	-.008 ***	-.011 ***			.003	.011 **
Effect p25 to p75 × Resources = Intermediate		-.003		-.001				-.014
Effect p25 to p75 × Resources = High		-.007		.003				.002
F (Kleibergen-Paap)					11.232	0.026		

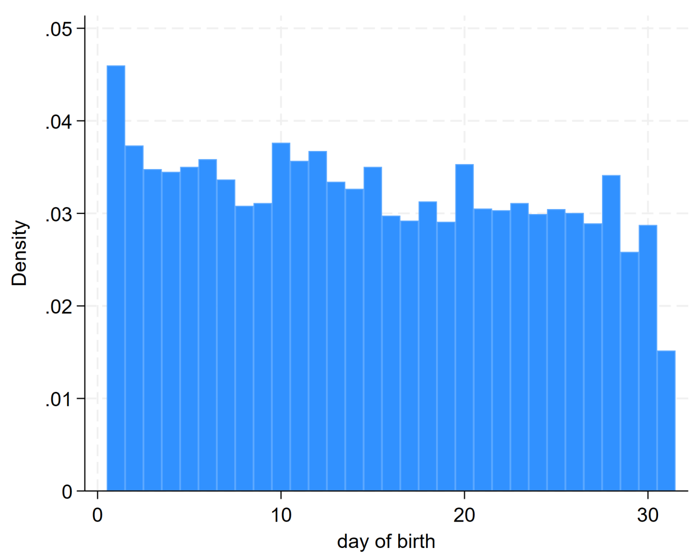


Figure A.1.3. Distribution of day of birth in the sample. The figure shows the density of reported days of birth (1–31) in the sample.

Table A.1.11. Second-stage results for subsequent facility birth using additive resources index. The dependent variable is a binary indicator equal to 1 if the next baby from the same mother was born in a healthcare facility. Resources is a composite index of facility characteristics, grouped into terciles. Controls include past and expected rainfall, the instrument measured at the time of the next birth, and their interactions with distance. Seasonality and time fixed effects are included. Standard errors are clustered at the mother level.

	All births		Last 5 years births					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Reduced form *Next baby Facility birth (0/1)*		**First stage** *Facility birth (0/1)*		**IV** *Next baby Facility birth (0/1)*		**Reduced form** *Next baby Facility birth (0/1)*	
Rain (m) × Distance (5km)	-0.393***	-0.490***	-0.397*	-0.391			-0.548***	-0.631***
	(0.128)	(0.115)	(0.241)	(0.252)			(0.203)	(0.212)
Rain (m) × Distance (5km) × Resources = Intermediate		0.413		0.227				0.148
		(0.426)		(0.937)				(0.805)
Rain (m) × Distance (5km) × Resources = High		0.482		0.413				0.540
		(0.376)		(1.259)				(0.917)
Facility birth (0/1)					1.380	1.746		
					(0.863)	(1.592)		
Facility birth (0/1) × Resources = Intermediate						0.317		
						(3.168)		
Facility birth (0/1) × Resources = High						-0.181		
						(2.638)		
N	8749	8749	2368	2368	2368	2368	2368	2368
N of clusters	7075	7075	2274	2274	2274	2274	2274	2274
Mean of Y	0.920	0.920	0.899	0.899	0.929	0.929	0.929	0.929
Effect p25 to p75	-.008 ***	-.01 ***	-.01 *	-.01			-.013 ***	-.015 ***
Effect p25 to p75 × Resources = Intermediate		-.002		-.004				-.012
Effect p25 to p75 × Resources = High		0		.001				-.002
F (Kleibergen-Paap)					2.722	0.180		

Table A.1.12. Sample characteristics by heaping day of birth. This table compares observable characteristics between births reported on heaping days (1st, 10th, 12th, 15th, 20th) and all other days.

	(1)	(2)	(3)
	Other days	Day of birth: 1, 10, 12, 15, 20	Difference (1)-(2)
Baby			
Birth order	2.94	2.98	-0.04**
	(1.93)	(1.97)	(0.02)
Infant mortality (0/1)	0.05	0.07	-0.02***
	(0.22)	(0.25)	(0.00)
Neonatal mortality (0/1)	0.02	0.03	-0.01***
	(0.15)	(0.17)	(0.00)
Received antenatal care (0/1)	0.98	0.98	0.00
	(0.13)	(0.13)	(0.00)
Received postnatal care (0/1)	0.45	0.43	0.02
	(0.50)	(0.50)	(0.01)
Mother			
Age	24.33	24.29	0.05
	(6.08)	(6.21)	(0.06)
Years of education	4.96	4.51	0.45***
	(3.68)	(3.68)	(0.04)
Total children ever born	4.81	5.06	-0.24***
	(2.31)	(2.36)	(0.02)
Visited health facility in last 12 months (0/1)	0.68	0.66	0.03***
	(0.47)	(0.48)	(0.00)
Village			
Rural (0/1)	0.84	0.85	-0.01***
	(0.37)	(0.36)	(0.00)
Distance to hospital (km)	6.84	6.96	-0.12*
	(6.69)	(6.57)	(0.06)
Closest facility has a doctor (0/1)	0.21	0.20	0.01*
	(0.41)	(0.40)	(0.00)
N	51,352	13,653	65,005

Table A.1.13. Estimates for neonatal mortality with heterogeneity by resources index. The dependent variable is a binary indicator equal to 1 if the baby died within 28 days of birth. Resources is a composite index of facility characteristics, grouped into terciles. Controls include past and expected rainfall, as well as their interactions with distance. Seasonality and time fixed effects are included. Standard errors are clustered at the grid-month level.

	All births		Last 5 years births					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Reduced form *Neonatal Mortality (0/1)*		**First stage** *Facility birth (0/1)*		**IV** *Neonatal Mortality (0/1)*		**Reduced form** *Neonatal Mortality (0/1)*	
Rain (m) × Distance (5km) × Resources = Low	-0.0147	-0.0153	-0.280***	-0.404***			-0.0219	-0.0394
	(0.0288)	(0.0372)	(0.0877)	(0.104)			(0.0382)	(0.0442)
Rain (m) × Distance (5km) × Resources = Intermediate		0.0161		0.511**				0.107
		(0.0744)		(0.236)				(0.0887)
Rain (m) × Distance (5km) × Resources = High		-0.0194		0.411				-0.110
		(0.0946)		(0.416)				(0.199)
Facility birth (0/1)					0.0784	0.0373		
					(0.143)	(0.195)		
Facility birth (0/1) × Resources = Intermediate						0.512		
						(1.808)		
Facility birth (0/1) × Resources = High						0.755		
						(1.220)		
N	46567	46567	12106	12106	12106	12106	12106	12106
N of clusters	811	811	748	748	748	748	748	748
Mean of Y	0.026	0.026	0.928	0.928	0.025	0.025	0.024	0.024
Effect p25 to p75	0	0	-.006 ***	-.009 ***			0	-.001
Effect p25 to p75 × Resources = Intermediate		0		.002				.001
Effect p25 to p75 × Resources = High		-.001		0				-.003
F (Kleibergen-Paap)					10.166	0.117		

Table A.1.14. Estimates for future healthcare visits with heterogeneity by facility staffing and baseline risk. The dependent variable is a binary indicator equal to 1 if the mother reported visiting a healthcare facility in the past 12 months. The outcome is measured at the mother level and observations are weighted by the inverse of the number of births per mother. High risk is a dummy equal to 1 if the standardized log-odds from a logistic regression of neonatal mortality on pre-birth covariates is above the median. Doctor equals 1 if the closest facility has at least one doctor on staff. Controls include past and expected rainfall, as well as their interactions with distance. Seasonality and time fixed effects are included. Standard errors are clustered at the mother level.

	All births		Last 5 years births					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Reduced form *Visited health facility last 12 months (0/1)*		**First stage** *Facility birth (0/1)*		**IV** *Visited health facility last 12 months (0/1)*		**Reduced form** *Visited health facility last 12 months (0/1)*	
Rain (m) × Distance (5km)	0.181	0.123	-0.372***	-0.350***			0.137	0.150
	(0.121)	(0.129)	(0.111)	(0.107)			(0.210)	(0.210)
Rain (m) × Distance (5km) × High risk = 1		0.143		-0.515				-0.374
		(0.226)		(0.408)				(0.594)
Rain (m) × Distance (5km) × Doctor = 1		-0.254		0.205				1.424
		(1.129)		(0.785)				(2.003)
Rain (m) × Distance (5km) × High risk = 1 × Doctor = 1		0.192		-0.0502				3.296
		(1.556)		(2.093)				(3.582)
Facility birth (0/1)					-0.370	-0.537		
					(0.542)	(1.010)		
Facility birth (0/1) × High risk = 1						0.297		
						(2.133)		
Facility birth (0/1) × Doctor = 1						1.733		
						(7.248)		
Facility birth (0/1) × High risk = 1 × Doctor = 1						-11.09		
						(29.14)		
N	43850	43850	9420	9420	9420	9420	9420	9420
N of clusters	12154	12154	7983	7983	7983	7983	7983	7983
Mean of Y	0.668	0.668	0.925	0.925	0.730	0.730	0.730	0.730
Effect p25 to p75		.003		-.008 ***				.003
Effect p25 to p75 × High risk = 1		.006		-.019 **				-.005
Effect p25 to p75 × Doctor = 1		-.003		-.003				.035
Effect p25 to p75 × High risk = 1 × Doctor = 1		.004		-.016				.101
F (Kleibergen-Paap)					11.232	0.027		

Table A.1.15. Estimates for subsequent facility birth with heterogeneity by facility staffing and baseline risk. The dependent variable is a binary indicator equal to 1 if the next baby from the same mother was born in a healthcare facility. High risk is a dummy equal to 1 if the standardized log-odds from a logistic regression of neonatal mortality on pre-birth covariates is above the median. Doctor equals 1 if the closest facility has at least one doctor on staff. Controls include past and expected rainfall, the instrument measured at the time of the next birth, and their interactions with distance. Seasonality and time fixed effects are included. Standard errors are clustered at the mother level.

	All births		Last 5 years births					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Reduced form		**First stage**		**IV**		**Reduced form**	
	Next baby				*Next baby*		*Next baby*	
	Facility birth (0/1)		*Facility birth (0/1)*		*Facility birth (0/1)*		*Facility birth (0/1)*	
Rain (m) × Distance (5km)	-0.393***	-0.440***	-0.397*	-0.382			-0.548***	-0.568***
	(0.128)	(0.107)	(0.241)	(0.242)			(0.203)	(0.208)
Rain (m) × Distance (5km) × High risk		0.523		-1.141				-0.231
		(0.336)		(0.857)				(0.597)
Rain (m) × Distance (5km) × Doctor		0.271		-4.867				-2.449
		(1.759)		(4.368)				(4.252)
Rain (m) × Distance (5km) × High risk × Doctor		-3.165		10.18*				5.978
		(2.254)		(5.299)				(5.441)
Facility birth (0/1)					1.380	1.177**		
					(0.863)	(0.537)		
Facility birth (0/1) × High risk						-0.788		
						(0.668)		
Facility birth (0/1) × Doctor						-0.567		
						(0.955)		
Facility birth (0/1) × High risk × Doctor						1.256		
						(3.032)		
N	8749	8749	2368	2368	2368	2368	2368	2368
N of clusters	7075	7075	2274	2274	2274	2274	2274	2274
Mean of Y	0.920	0.920	0.899	0.899	0.929	0.929	0.929	0.929
Effect p25 to p75	-.008 ***	-.009 ***	-.01 *	-.009			-.013 ***	-.014 ***
Effect p25 to p75 × High risk = 1		.002		-.037 *				-.019
Effect p25 to p75 × Doctor = 1		-.004		-.128				-.074
Effect p25 to p75 × High risk = 1 × Doctor = 1		-.06 *		.092				.067
F (Kleibergen-Paap)					2.722	0.805		

Table A.1.16. Two-sample 2SLS estimates with heterogeneity by facility staffing and baseline risk. The first stage is estimated in the subsample of births within the past five years, and the estimated first-stage coefficients are applied to the full sample to generate predicted values of the endogenous regressor. The second stage is then run in the full sample using these predicted values. High risk equals 1 if the standardized log-odds from a logistic regression of neonatal mortality on pre-birth covariates is above the median. Doctor equals 1 if the closest facility has at least one doctor on staff. Controls include past and expected rainfall, the instrument measured at the time of the next birth, and their interactions with distance. Seasonality and time fixed effects are included. Standard errors are clustered at the mother level.

	(1)	(2)	(3)	(4)	(5)	(6)
	Neonatal mortality (0/1)		*Visited health facility last 12 months (0/1)*		*Next baby facility birth (0/1)*	
Facility (0/1)	0.053	0.258	-0.488	-0.386	1.012	0.737
	(0.208)	(0.604)	(8.920)	(2.754)	(1.068)	(3.796)
Facility (0/1) × High risk = 1		-0.244		0.110		-1.068
		(0.511)		(3.905)		(4.976)
Facility (0/1) × Doctor = 1		-0.424		-0.251		-1.710
		(2.460)		(22.694)		(4.257)
Facility (0/1) × Doctor = 1 × High risk = 1		-0.278		0.486		0.263
		(2.518)		(12.806)		(6.937)

Table A.1.17. Two-sample 2SLS estimates with heterogeneity by resources index. The first stage is estimated in the subsample of births within the past five years, and the estimated first-stage coefficients are applied to the full sample to generate predicted values of the endogenous regressor. The second stage is then run in the full sample using these predicted values. Resources is a composite index of facility characteristics, grouped into terciles. Controls include past and expected rainfall, the instrument measured at the time of the next birth, and their interactions with distance. Seasonality and time fixed effects are included. Standard errors are clustered at the mother level.

	(1)	(2)	(3)	(4)	(5)	(6)
	Neonatal mortality (0/1)		*Visited health facility last 12 months (0/1)*		*Next baby facility birth (0/1)*	
Facility (0/1)	0.053	0.041	-0.488	-0.866	1.012	1.137
	(0.208)	(0.185)	(8.920)	(3.643)	(1.068)	(4.226)
Facility (0/1) × Resources = Intermediate		-0.094		-1.640		-0.388
		(0.656)		(10.255)		(6.171)
Facility (0/1) × Resources = High		0.110		1.826		-0.048
		(0.894)		(11.771)		(2.239)

A.2 Chapter 2 Appendix

Table A.2.1. Impact of colleges on take-up of healthcare by residence location. * $p<0.10$, ** $p<0.05$, *** $p<0.01$. Standard errors clustered at the district level. District as well as survey-year fixed effects used. The coefficients estimated are the parameters in equation 2.1. Variables are available in all datasets, viz. 4 rounds of DHS and 3 rounds of DLHS. 'Any ANC' variable is reported as a percentage. 'ANC visits' denotes the number of visits, 'First ANC' denotes the first month of ANC, for those who had an ANC visit. Variables available for the last pregnancy.

	ANC Visits		First ANC		Any ANC	
	(1)	(2)	(3)	(4)	(5)	(6)
	Rural	Urban	Rural	Urban	Rural	Urban
Graduate batches	0.002	0.018***	-0.004	-0.001	0.000	-0.000
	(0.005)	(0.006)	(0.004)	(0.002)	(0.000)	(0.000)
College opening	0.122**	-0.031	0.025	0.028	0.002	-0.000
	(0.058)	(0.054)	(0.034)	(0.026)	(0.002)	(0.001)
Spillover	0.034	0.160**	0.063***	0.053*	-0.002	-0.000
	(0.052)	(0.068)	(0.020)	(0.030)	(0.001)	(0.001)
Mean Dep. Var.	3.389	4.627	2.888	2.601	0.673	0.676
Observations	587704	181871	530678	175087	2656950	778683

Table A.2.2. Impact of colleges on share of ANC visits by provider type and residence location. * $p<0.10$, ** $p<0.05$, *** $p<0.01$. Standard errors clustered at the district level. District as well as survey-year fixed effects used. The coefficients estimated are the parameters in equation 2.1. All variables are reported as percentages. ANC_none is available in all datasets. ANC providers are present in DHS datasets and DLHS 2003. Within these, Anganwadi and ASHA are only present in DHS. ANC provider questions are asked to those who saw someone for an ANC. Categories of 'village health worker' and 'other' ANC provider not reported here, since their mean shares are merely 0.63% and 0.19% respectively.

	Doctor		Nurse		ASHA		TBA		Anganwadi		None	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban
Graduate batches	0.204**	0.106	-0.216***	0.012	-0.742***	-0.471***	-0.009	-0.005	-0.521***	-0.204	0.099*	0.070*
	(0.086)	(0.103)	(0.072)	(0.052)	(0.145)	(0.078)	(0.006)	(0.004)	(0.176)	(0.125)	(0.057)	(0.041)
College opening	3.057***	2.457***	1.081	-1.217	-0.308	0.435	-0.128	0.049	-0.730	-0.024	0.562	-0.387
	(0.888)	(0.870)	(1.006)	(0.936)	(0.741)	(0.846)	(0.095)	(0.066)	(0.825)	(0.794)	(0.796)	(0.436)
Spillover	0.135	1.101**	-1.203**	-0.951*	-1.632	-2.012	-0.010	0.027	0.044	-2.778*	-0.608	-0.251
	(0.520)	(0.507)	(0.602)	(0.543)	(1.409)	(1.495)	(0.043)	(0.055)	(1.550)	(1.523)	(0.482)	(0.303)
Mean Dep. Var.	34.153	45.031	27.992	20.810	14.166	6.567	0.828	0.508	16.102	10.325	20.643	8.365
Observations	590917	204514	590529	204270	257143	70022	386559	121878	257143	70022	772261	257653

Table A.2.3. Impact of colleges on share of deliveries assisted by provider type and residence location. * $p<0.10$, ** $p<0.05$, *** $p<0.01$. Standard errors clustered at the district level. District as well as survey-year fixed effects used. The coefficients estimated are the parameters in equation 2.1. All variables are reported as percentages. Variables are available for women who have had at least one birth. Delivery provider variables are defined for the last birth in DLHS and for available births in DHS. Delivery by doctor, nurse, TBA, none, or other are available in all datasets. Skilled birth denotes if the birth was assisted by either a doctor, nurse or a trained health professional. Here, TBA means 'traditional birth attendant'. The category of 'other health professional' not reported here since its mean share is merely 1.7%. Facility denotes whether the birth took place in any medical facility, and not at home/on the way to a facility.

	Facility		Skilled Birth		Doctor		Nurse		TBA		Other		None	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban
Graduate batches	0.103	-0.133***	-0.216***	-0.232***	0.218**	0.074	-0.396***	-0.232	0.077	0.180***	-0.018	-0.013	0.016	-0.007
	(0.074)	(0.049)	(0.069)	(0.082)	(0.088)	(0.067)	(0.081)	(0.195)	(0.094)	(0.060)	(0.027)	(0.018)	(0.010)	(0.008)
College opening	0.735	1.164**	-2.146***	-1.191*	1.738*	0.850	-2.905***	-1.785**	-0.188	-0.122	0.071	-0.016	0.004	0.045
	(0.741)	(0.570)	(0.706)	(0.667)	(1.031)	(0.755)	(0.879)	(0.883)	(0.888)	(0.801)	(0.216)	(0.132)	(0.082)	(0.065)
Spillover	-0.180	-1.150***	-2.083***	-1.859**	-0.258	-0.279	-2.738***	-2.969**	-1.207*	-0.521	1.040***	0.643***	0.183***	0.150**
	(0.509)	(0.413)	(0.564)	(0.811)	(0.598)	(0.740)	(0.822)	(1.194)	(0.724)	(0.705)	(0.226)	(0.168)	(0.067)	(0.060)
Mean Dep. Var.	45.587	56.252	51.377	71.507	27.223	42.322	33.676	41.041	27.608	15.875	5.715	3.276	0.692	0.390
Observations	889393	277099	617720	134077	683709	158359	683709	158359	683709	158359	671664	155709	683709	158359

Table A.2.4. Impact of colleges on health status by residence location. * $p<0.10$, ** $p<0.05$, *** $p<0.01$. Standard errors clustered at the district level. District as well as survey-year fixed effects used. The coefficients estimated are the parameters in equation 2.1. Morbidity variables reported as percentages, mortality variables reported per thousand. Mortality variables available in all datasets, viz. four rounds of DHS and three rounds of DLHS. Restricted to 5 years before the interview year for mortality variables. Morbidity variables only in DHS, with reference year as the current interview year. Given that only one kind of data (i.e., only DHS) is used for morbidity, dataset fixed effects not used. Neonatal, infant and under-5 mortality refer to death within one month, one year and 5 years of birth, respectively.

	Diarrhoea		Fever		Cough		Neonatal		Infant		Under-5	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban
Graduate batches	-0.127**	-0.171**	-0.160**	-0.058	-0.180**	-0.096	-0.165	0.187	0.046	0.384	0.194	0.490*
	(0.064)	(0.074)	(0.072)	(0.081)	(0.078)	(0.096)	(0.152)	(0.225)	(0.143)	(0.277)	(0.140)	(0.289)
College opening	0.158	-0.045	0.523	0.039	0.535	0.182	-0.426	-0.340	0.943	0.199	2.656	0.530
	(0.742)	(0.626)	(0.922)	(0.803)	(0.787)	(0.828)	(1.160)	(1.488)	(1.610)	(1.809)	(1.820)	(2.130)
Spillover	-0.757***	-0.209	-0.523	-0.633	-0.498	-0.848	0.814	0.957	1.723*	0.366	3.356***	0.026
	(0.273)	(0.409)	(0.378)	(0.526)	(0.376)	(0.617)	(0.878)	(1.323)	(1.019)	(1.576)	(1.126)	(1.863)
Mean Dep. Var.	8.932	8.353	14.310	13.980	14.027	13.789	31.378	22.893	45.151	33.000	52.581	37.268
Observations	328672	75765	328759	75807	328738	75792	863275	225877	863275	225877	863275	225877

A.3 Chapter 3 Appendix

Table A.3.1. Summary statistics for the analysis sample. The sample includes all births in the DHS surveys from 38 Sub-Saharan African countries for which place of delivery is recorded. *Facility delivery*, *Hospital delivery*, and *Lower-level facility delivery* are indicators for the place of birth. *Day-of-birth death* is an indicator equal to one if the child died on the day of birth. The *risk score* sums six indicators of mother-side risk factors as defined in Section 3.3.1.

	mean	sd	min	max	count
Facility delivery (0/1)	0.70	0.46	0	1	317939
Hospital delivery (0/1)	0.52	0.50	0	1	317939
Lower-level facility delivery (0/1)	0.19	0.39	0	1	317939
Day-of-birth death (0/1)	0.01	0.09	0	1	317939
Neonatal death (0/1)	0.03	0.16	0	1	317935
Risk score (0–6)	1.07	0.97	0	6	317939
High risk (risk score ≥ 2)	0.30	0.46	0	1	317939
High parity (birth order ≥ 5)	0.29	0.45	0	1	317939
Twin birth (0/1)	0.04	0.19	0	1	317939
Maternal age <20 or ≥ 40 (0/1)	0.21	0.41	0	1	317939
No formal education (0/1)	0.38	0.48	0	1	317939
Birth interval $<$ 18 months (0/1)	0.05	0.22	0	1	317939
Prior miscarriage (0/1)	0.01	0.11	0	1	317724

Table A.3.2. Village-level summary statistics. Each observation is a village cluster in the analysis sample. *Births per village* is the number of births recorded for the village across the five-year DHS observation window. *Facility births per village* and *Day-of-birth deaths per village* are counts; *Share facility delivery*, *Village with $\geq$ 1 day-of-birth death*, *Mean risk score*, and *Share high-risk pregnancies* are village-level averages.

	mean	sd	min	max	count
Births per village	15.08	9.64	1	160	21079
Facility births per village	10.63	7.77	0	129	21079
Share facility delivery	0.77	0.30	0	1	21079
Day-of-birth deaths per village	0.13	0.41	0	6	21079
Village with $\geq$ 1 day-of-birth death	0.10	0.31	0	1	21079
Mean risk score	0.90	0.54	0	5	21079
Share high-risk pregnancies	0.24	0.22	0	1	21079

Table A.3.3. Village-year summary statistics. Each observation is a village-year cell within the five-year DHS observation window preceding each country's survey. *Births per village-year* is the total number of births recorded in the village in that year. *Facility births* and *Home births per village-year* are the corresponding counts by place of delivery. *Share facility delivery* is the proportion of births in the village-year that occurred in a healthcare facility. *Day-of-birth deaths per village-year* is the total number of deaths occurring on the day of birth, with separate counts for deaths following facility and home births. *Day-of-birth mortality rate, facility births* is the share of facility births that resulted in a day-of-birth death; the analogous variable is defined for home births. Rates are set to zero in village-years with no births of the corresponding type. *Village-year with* $\geq$ *1 day-of-birth death* is an indicator equal to one if any day-of-birth death occurred in the village in that year.

	mean	sd	min	max	count
Births per village-year	3.65	2.59	1	40	87147
Facility births per village-year	2.57	2.18	0	35	87147
Home births per village-year	1.08	1.92	0	23	87147
Share facility delivery	0.75	0.36	0	1	87147
Day-of-birth deaths per village-year	0.03	0.19	0	4	87147
Day-of-birth deaths in facility per village-year	0.02	0.16	0	3	87147
Day-of-birth deaths at home per village-year	0.01	0.09	0	4	87147
Day-of-birth mortality rate, facility births	0.01	0.06	0	1	87147
Day-of-birth mortality rate, home births	0.00	0.05	0	1	87147
Village-year with $\geq$ 1 day-of-birth death	0.03	0.17	0	1	87147

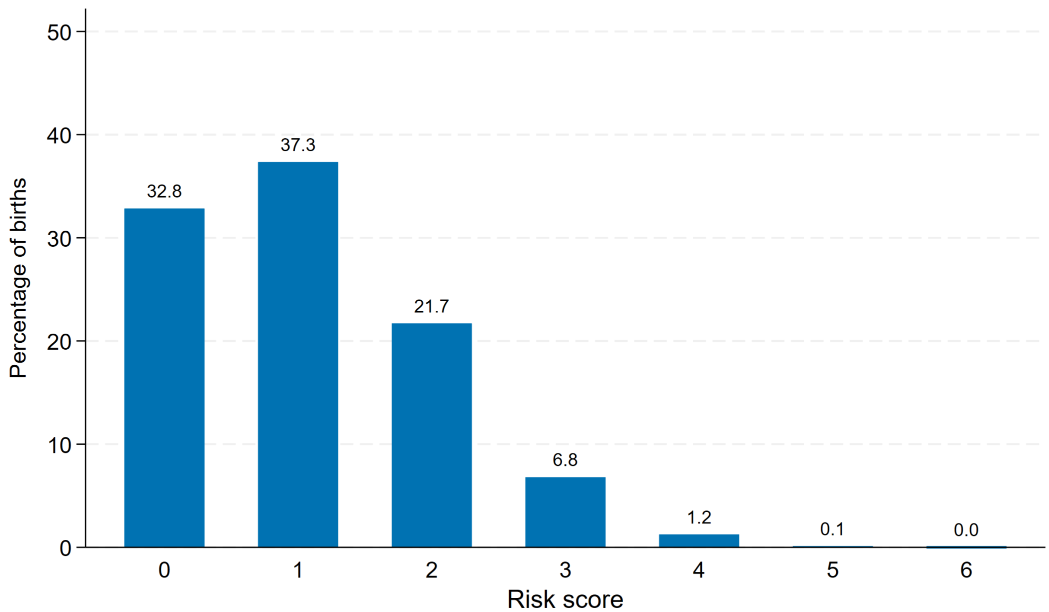


Figure A.3.1. Distribution of the risk score across births in the analysis sample. The risk score sums six binary indicators of mother-side risk factors: high parity (birth order ≥ 5), twin birth, maternal age below 20 or 40 and above, no formal education, a preceding birth interval shorter than 18 months, and a prior miscarriage.

Bibliography

Adhvaryu, Achyuta, "Learning, misallocation, and technology adoption: evidence from new malaria therapy in Tanzania," *The Review of economic studies*, 2014, *81* (4), 1331–1365.

— **and Anant Nyshadham**, "Return to treatment in the formal health care sector: Evidence from Tanzania," *American Economic Journal: Economic Policy*, 2015, *7* (3), 29–57.

Adhvaryu, Achyuta R. and Anant Nyshadham, "Healthcare Choices, Information and Health Outcomes," September 2010.

Agte, Patrick and Jitendra Kumar Soni, "Fighting Silent Killers: How India's Public Healthcare Staffing Expansion Saves Lives by Improving Access and Market Quality," *Working paper.*

Allcott, Hunt and Todd Rogers, "The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation," *American Economic Review*, 2014, *104* (10), 3003–3037.

Andegiorgish, Amanuel Kidane, Zhonghai Zhu, Mohamed Elhoumed, Qi Qi, and Lingxia Zeng, "Determinants of institutional delivery in Kenya. Evidence from population and Demographic health survey,2014," preprint, In Review May 2021.

Andrew, Alison and Marcos Vera-Hernández, "Incentivizing demand for supply-constrained care: Institutional birth in India," *Review of Economics and Statistics*, 2022, pp. 1–45.

Arntz, Melanie, Eduard Brüll, and Cäcilia Lipowski, "Do preferences for urban amenities differ by skill?," *Journal of Economic Geography*, May 2023, *23* (3), 541–576.

Bandiera, Oriana and Imran Rasul, "Social networks and technology adoption in northern Mozambique," *The economic journal*, 2006, *116* (514), 869–902.

Banerjee, Abhijit Vinayak, Esther Duflo, Rachel Glennerster, and Dhruva Kothari, "Improving immunisation coverage in rural India: clustered randomised controlled evaluation of immunisation campaigns with and without incentives," *Bmj*, 2010, *340.*

Barron, Kai, Steffen Huck, and Philippe Jehiel, "Everyday econometricians: Selection neglect and overoptimism when learning from others," *American Economic Journal: Microeconomics*, 2024, *16* (3), 162–198.

Becker, Gary S and Kevin M Murphy, "A theory of rational addiction," *Journal of political Economy*, 1988, *96* (4), 675–700.

Bove, Frank J, Mark C Fulcomer, Judith B Klotz, Jorge Esmart, Ellen M Dufficy, and Jonathan E Savrin, "Public drinking water contamination and birth outcomes," *American Journal of Epidemiology*, 1995, *141* (9), 850–862.

Bove, Frank, Youn Shim, and Perri Zeitz, "Drinking water contaminants and adverse pregnancy outcomes: a review.," *Environmental health perspectives*, 2002, *110* (suppl 1), 61–74.

Budillon, Fulvia, "Beyond Access: Facility Birth, Healthcare Use, and Early-Life Mortality in Malawi," forthcoming. Working paper.

Bursztyn, Leonardo, Florian Ederer, Bruno Ferman, and Noam Yuchtman, "Understanding Mechanisms Underlying Peer Effects: Evidence from a Field Experiment on Financial Decisions," *Econometrica*, 2014, *82* (4), 1273–1301.

Callaway, Brantly and Pedro HC Sant'Anna, "Difference-in-differences with multiple time periods," *Journal of Econometrics*, 2021, *225* (2), 200–230.

Chandra, Amitabh, Amy Finkelstein, Adam Sacarny, and Chad Syverson, "Health care exceptionalism? Performance and allocation in the US health care sector," *American Economic Review*, 2016, *106* (8), 2110–2144.

Chaudhury, Nazmul, Jeffrey Hammer, Michael Kremer, Karthik Muralidharan, and F. Halsey Rogers, "Missing in Action: Teacher and Health Worker Absence in Developing Countries," *Journal of Economic Perspectives*, February 2006, *20* (1), 91–116.

Che, Yi and Lei Zhang, "Human Capital, Technology Adoption and Firm Performance: Impacts of China's Higher Education Expansion in the Late 1990s," *The Economic Journal*, 2018, *128* (614), 2282–2320. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ecoj.12524.

Conley, Timothy G and Christopher R Udry, "Learning about a new technology: Pineapple in Ghana," *American economic review*, 2010, *100* (1), 35–69.

Das, Jishnu, Benjamin Daniels, Monisha Ashok, Eun-Young Shim, and Karthik Muralidharan, "Two Indias: The structure of primary health care markets in rural Indian villages with implications for policy," *Social Science & Medicine (1982)*, May 2022, *301*, 112799.

Diamond, Rebecca, "The Determinants and Welfare Implications of US Workers' Diverging

Location Choices by Skill: 1980-2000," *The American Economic Review*, 2016, *106* (3), 479–524. Publisher: American Economic Association.

Dingel, Jonathan I, Joshua D Gottlieb, Maya Lozinski, and Pauline Mourot, "Market Size and Trade in Medical Services," *Working Paper.*

Dranove, David, Daniel Kessler, Mark McClellan, and Mark Satterthwaite, "Is more information better? The effects of "report cards" on health care providers," *Journal of political Economy*, 2003, *111* (3), 555–588.

Dupas, P. and E. Miguel, "Chapter 1 - Impacts and Determinants of Health Levels in Low-Income Countries," in Abhijit Vinayak Banerjee and Esther Duflo, eds., *Handbook of Economic Field Experiments*, Vol. 2 of *Handbook of Economic Field Experiments*, North-Holland, January 2017, pp. 3–93.

Dupas, Pascaline et al., "Health behavior in developing countries," *Annual review of Economics*, 2011, *3* (1), 425–449.

Esponda, Ignacio, "Behavioral equilibrium in economies with adverse selection," *American Economic Review*, 2008, *98* (4), 1269–1291.

_ **and Emanuel Vespa**, "Endogenous sample selection: A laboratory study," *Quantitative Economics*, 2018, *9* (1), 183–216.

Foster, Andrew D and Mark R Rosenzweig, "Learning by doing and learning from others: Human capital and technical change in agriculture," *Journal of political Economy*, 1995, *103* (6), 1176–1209.

Frenette, Marc, "Too Far to Go On? Distance to School and University Participation," *Education Economics*, March 2006. Publisher: Routledge.

_ , "Do universities benefit local youth? Evidence from the creation of new universities," *Economics of Education Review*, June 2009, *28* (3), 318–328.

Fujiwara, Thomas, Kyle Meng, and Tom Vogl, "Habit formation in voting: Evidence from rainy elections," *American Economic Journal: Applied Economics*, 2016, *8* (4), 160–188.

Gabrysch, Sabine, Robin C Nesbitt, Anja Schoeps, Lisa Hurt, Seyi Soremekun, Karen Edmond, Alexander Manu, Terhi J Lohela, Samuel Danso, Keith Tomlin et al., "Does facility birth reduce maternal and perinatal mortality in Brong Ahafo, Ghana? A secondary analysis using data on 119 244 pregnancies from two cluster-randomised controlled trials," *The Lancet Global Health*, 2019, *7* (8), e1074–e1087.

Geweke, John, Gautam Gowrisankaran, and Robert J Town, "Bayesian inference for hospital quality in a selection model," *Econometrica*, 2003, *71* (4), 1215–1238.

Gowrisankaran, Gautam and Robert J. Town, "Estimating the quality of care in hospitals using instrumental variables," *Journal of Health Economics*, December 1999, *18* (6), 747–767.

Haye, Sidra, "Effect of patient death on referrals to cardiac specialists," *Health Economics*, 2024, *33* (8), 1857–1868.

Hindu, The, "Incentives for doctors serving in remote areas of Odisha," *The Hindu*, April 2015.

Holte, Jon Helgheim, Trine Kjaer, Birgit Abelsen, and Jan Abel Olsen, "The impact of pecuniary and non-pecuniary incentives for attracting young doctors to rural general practice," *Social Science & Medicine*, March 2015, *128*, 1–9.

Hu, Dan, Baisong Zhang, Mingyu Huang, Min Liu, Xiulong Xia, Yanli Zuo, and Xiaoyun Liu, "Evaluation of a medical education policy with compulsory rural service in China," *Frontiers in Public Health*, February 2023, *11*, 1042898.

Hussam, Reshmaan, Atonu Rabbani, Giovanni Reggiani, and Natalia Rigol, "Rational habit formation: experimental evidence from handwashing in India," *American Economic Journal: Applied Economics*, 2022, *14* (1), 1–41.

Jagnani, Maulik and Gaurav Khanna, "The effects of elite public colleges on primary and secondary schooling markets in India," *Journal of Development Economics*, September 2020, *146*, 102512.

Jeworrek, Sabrina and Matthias Brachert, "Where to go? High-skilled individuals' regional preferences," *IWH Discussion Paper*, 2022.

Jones, Damon, David Molitor, and Julian Reif, "Incentives and Habit Formation in Health Screenings: Evidence from the Illinois Workplace Wellness Study," Technical Report, National Bureau of Economic Research 2024.

Jr, Joseph J Doyle, "Health insurance, treatment and outcomes: using auto accidents as health shocks," *Review of Economics and Statistics*, 2005, *87* (2), 256–270.

Jung, Dawoon and Booyuel Kim, "The effect of health facility births on newborn mortality in Malawi," *International Journal of Health Economics and Management*, 2024, *24* (3), 393–406.

Kruk, Margaret E., Peter C. Rockers, Godfrey Mbaruku, Magdalena M. Paczkowski, and Sandro Galea, "Community and health system factors associated with facility delivery in rural Tanzania: A multilevel analysis," *Health Policy*, October 2010, *97* (2), 209–216.

Lapid, Patrick A, "Expanding College Access: The Impact of New Universities on Local

Enrolments," *Job Market Paper, University of California, Berkeley*, 2018.

Leonard, Kenneth L, "Learning in health care: evidence of learning about clinician quality in Tanzania," *Economic Development and Cultural Change*, 2007, *55* (3), 531–555.

Leslie, Hannah H, Günther Fink, Humphreys Nsona, and Margaret E Kruk, "Obstetric facility quality and newborn mortality in Malawi: a cross-sectional study," *PLoS medicine*, 2016, *13* (10), e1002151.

Lowes, Sara and Eduardo Montero, "The legacy of colonial medicine in Central Africa," *American Economic Review*, 2021, *111* (4), 1284–1314.

Mahal, Ajay and Manoj Mohanan, "Growth of private medical education in India," *Medical Education*, October 2006, *40* (10), 1009–1011.

Manski, Charles F, "Identification of endogenous social effects: The reflection problem," *The review of economic studies*, 1993, *60* (3), 531–542.

Manthalu, Gerald, Deokhee Yi, Shelley Farrar, and Dominic Nkhoma, "The effect of user fee exemption on the utilization of maternal health care at mission health facilities in Malawi," *Health policy and planning*, 2016, *31* (9), 1184–1192.

Muralidharan, Karthik, *Accelerating India's Development* 2024.

Newhouse, Joseph P., Albert P. Williams, Bruce W. Bennett, and William B. Schwartz, "Does the Geographical Distribution of Physicians Reflect Market Failure?," *The Bell Journal of Economics*, 1982, *13* (2), 493–505. Publisher: [RAND Corporation, Wiley].

of India, Govt, "Medical Infrastructure in India," July 2018.

Okeke, Edward N, "When a doctor falls from the sky: The impact of easing doctor supply constraints on mortality," *American economic review*, 2023, *113* (3), 585–627.

Okeke, Edward N., "When a Doctor Falls from the Sky: The Impact of Easing Doctor Supply Constraints on Mortality," *American Economic Review*, March 2023, *113* (3), 585–627.

Okeke, Edward N, Amalavoyal V Chari et al., "Can institutional deliveries reduce newborn mortality? Evidence from Rwanda," *Citeseer.[Google Scholar]*, 2015.

Pacific, The Lancet Regional Health – Western, "Junior doctor strikes in South Korea: more doctors are needed?," *The Lancet Regional Health: Western Pacific*, March 2024, *44*, 101056.

Rogall, Thorsten, "Mobilizing the masses for genocide," *American Economic Review*, 2021, *111* (1), 41–72.

Rosenthal, Meredith B, Alan Zaslavsky, and Joseph P Newhouse, "The Geographic Distribution of Physicians Revisited," *Health Services Research*, December 2005, *40* (6 Pt 1), 1931–1952.

Sarsons, Heather, "Interpreting signals in the labor market: evidence from medical referrals," *Job market paper*, 2017, pp. 141–145.

Scott, Anthony, Julia Witt, John Humphreys, Catherine Joyce, Guyonne Kalb, Sung-Hee Jeon, and Matthew McGrail, "Getting doctors into the bush: General Practitioners' preferences for rural location," *Social Science & Medicine*, November 2013, *96*, 33–44.

Shajarizadeh, Ali and Karen Ann Grépin, "The impact of institutional delivery on neonatal and maternal health outcomes: evidence from a road upgrade programme in India," *BMJ Global Health*, July 2022, *7* (7), e007926.

Singh, Manasvini, "Heuristics in the delivery room," *Science*, 2021, *374* (6565), 324–329.

Swami, Megha and Anthony Scott, "Impact of rural workforce incentives on access to GP services in underserved areas: Evidence from a natural experiment," *Social Science & Medicine*, July 2021, *281*, 114045.

Thompson, Emmanuel and Seidu Sofo, "Maternal Mortality in Ghana: An Econometric Analysis," January 2014, *3*, 115–124.

Tura, Gurmesa, Mesganaw Fantahun, and Alemayehu Worku, "The effect of health facility delivery on neonatal mortality: systematic review and meta-analysis," *BMC pregnancy and childbirth*, 2013, *13* (1), 18.

Vu, Khoa and Tu-Anh Vu-Thanh, "Higher Education Expansion, Labor Market, and Firm Productivity in Vietnam," *Working Paper*, 2024.

Weaver, Jeffrey, Sandip Sukhtankar, Paul Niehaus, and Karthik Muralidharan, "Cash Transfers for Child Development: Experimental Evidence from India," January 2024.

WHO, "Newborn Mortality," Fact sheet 2024.

Worku, Abebaw Gebeyehu, Alemayehu Worku Yalew, and Mesganaw Fantahun Afework, "Maternal complications and women's behavior in seeking care from skilled providers in North Gondar, Ethiopia," *PloS one*, 2013, *8* (3), e60171.

World Bank, "Mortality rate, under-5 (per 1,000 live births)," https://data.worldbank.org/indicator/SH.DYN.MORT?locations=ZG 2023. Accessed: 2025-07-24.

_ , "World Development Indicators," 2025. Accessed July 2025.