Saving Neonatal Lives for a Quarter

Christine Valente, Hans H. Sievertsen and Mahesh C. Puri

Discussion Paper 20/728

September 2021

School of Economics

University of Bristol
Priory Road Complex
Bristol
BS8 1TU
United Kingdom

University of BRISTOL
School of Economics

Saving Neonatal Lives for a Quarter

Christine Valente (University of Bristol and IZA)*
ⓡ Hans H. Sievertsen (University of Bristol, VIVE, and IZA)
ⓡ Mahesh C. Puri (Center for Research on Environment, Health and Population Activities)

September 2021

Abstract
Neonatal sepsis kills over 400,000 children annually. Experimental findings on the preventive use of chlorhexidine vary widely across settings, leading to external validity concerns. To address this, we (i) provide the first quasi-experimental estimates of the effect of chlorhexidine in "real life" conditions and (ii) apply machine-learning (ML) to analyze treatment effect heterogeneity in a nationally-representative, Nepalese observational dataset. We find that chlorhexidine decreases neonatal mortality by 43% and that a simple targeting policy leveraging heterogeneous treatment effects improves neonatal survival relative to WHO recommendations. Out-of-sample ML predictions match the heterogeneous pattern of existing experimental findings across five different countries.

Keywords: neonatal mortality, chlorhexidine, Nepal
JEL Classification: I18, J13, O15

*Corresponding author: christine.valente@bristol.ac.uk. School of Economics, University of Bristol, Priory Road Complex, Priory Road, Bristol, BS8 1TU, UK. We thank Leela Khanal for sharing implementation details of the Chlorhexidine Navi(Cord) Care Program (CHX-NCP). For their useful comments and suggestions, we thank Douglas Almond, Clément de Chaisemartin, Thomas Dee, Lena Edlund, Mette Ejrnæs, Anna Folke Larsen, Xavier d'Haultfœuille, Stephanie von Hinke, Ulrik Hvidman, Grant Miller, Helen Simpson, Stéphanie Vincent Lyk-Jensen, Miriam Wüst, and presentation participants at Royal Holloway, PAA Annual Meeting 2021, Essen Health Conference 2021 and IAAE Conference 2021. Funding from the University of Bristol's Global Challenges Research Fund is gratefully acknowledged.

1 Introduction

Contrary to a commonly held belief, prevention is not always desirable (Kowalski, 2021; Newhouse, 2021). In particular, prevention is unlikely to benefit everyone, so that recommendations regarding whether to- and who should use preventive measures evolve along with scientific evidence which may appear contradictory (World Health Organization, 2020; Kowalski, 2021).

In this paper we study the prevention of a common cause of neonatal mortality in low-income settings. Each year, it is estimated that as many as 2.7 Million children die within the first 28 days of their lives, of which 400,000 die of blood infection (or "neonatal sepsis") (Liu et al., 2016). Starting fifteen years ago, three randomized controlled trials (RCT) brought hopes that the simple preventive application of the disinfectant chlorhexidine (CHX) to the umbilical cord stump may eradicate this condition (see Mullany et al., 2006; El Arifeen et al., 2012; Soofi et al., 2012, who report a 20-38% decrease in neonatal mortality). Replication however failed in two further trials, leading experts to express doubts about the effectiveness of CHX application at scale (Semrau et al., 2016; Sazawal et al., 2016; Osrin and Colbourn, 2016; Ponce Hardy, 2018).

The trials finding significant beneficial effects of CHX took place in localized areas of Nepal, Bangladesh, and Pakistan where nearly all births occur at home, whereas both trials finding no effect were carried out in Tanzanian and Zambian communities where neonatal mortality is lower and most births occur in health facilities. As a consequence, since 2013 the World Health Organization (WHO) recommends the application of CHX for home births only, in settings with neonatal mortality above 30 per 1000, and to otherwise stick to the default advice of keeping the cord stump clean and dry (World Health Organization, 2015). While a CHX dose costs as little as US$0.23 per dose (Hodgins et al., 2013), the WHO recommendation of restricting the use of CHX can be understood as balancing proven benefits against broader costs. These include the risk of application of other, potentially harmful substances, as well as the risk of diverting human, logistical, and financial resources away from other essential medicines and tasks in an area where the gap between recommended health care and practice is already large.[1]

[1]Friberg et al. (2010) and Requejo et al. (2015) document important gaps in the implementation of health practices preventing early life mortality in low-income countries. In the case of

This paper answers three open questions: (i) What is the effect of CHX in "real-life" conditions?; (ii) What drives the heterogeneity of RCT findings across different settings?; and (iii) Can an alternative targeting policy to the current WHO recommendation further reduce neonatal mortality?

Our first contribution is to provide the first estimates of the effect of CHX outside an experimental setting, which we do in a nationally representative sample for Nepal. Concerns about the scalability of experimental findings typically emphasize factors which lead to *smaller* treatment effects at scale — such as selected and non-representative samples, high compliance and adherence to protocol which cannot be replicated in "real-world" conditions (Al-Ubaydli et al., 2017). But in the case of trials of preventive health care measures such as CHX cord care, the treatment effect might be muted in ethical trials for two main reasons. First, subjects involved in these trials are referred to the hospital where needed — e.g., if signs of cord infection (omphalitis) appear during the frequent research team visits. This improved remedial care would tend to reduce the estimated effect of prevention on mortality. Second, there may not be a pure control group. In the case of CHX trials, both treated and control groups typically receive a comprehensive package of measures preventing omphalitis, which again reduces the relevance of CHX application (Semrau et al., 2016) and go well beyond the usual standard of care in low-income settings.[2,3] Indeed, in both trials finding no significant effect on mortality, the authors note that the neonatal mortality rate (NMR) was between 32 percent and 40 percent lower than in the most recent Demographic and Health Survey for the relevant area — even in the control group. Many factors may therefore lead to differences in CHX- and other preventive treatment effects in- and outside an experimental setting, in a direction that is unclear *a priori*.

Our second contribution is to apply machine-learning (ML) techniques recently developed by Athey et al. (2019) and Athey and Wager (2021) for use

CHX specifically, CHX cord care is still only at varying stages of implementation in 11 countries (PATH, 2017) even though many more countries have neonatal mortality rates above the WHO recommendation threshold in at least some regions (WHO, 2020).

[2]The typical package of services received by both control and treated subjects in CHX application trials are: a clean delivery kit, referral to clinic in the presence of danger signs, newborn health messages, antenatal clinic visits, and home visits starting soon after birth.

[3]Another potential channel through which CHX cord care may prevent neonatal death is by preventing neonatal tetanus (Bennett et al., 1997). In the Nepal CHX trial, this potential pathway to impact was also shut down by ensuring full maternal tetanus immunization at enrolment in the trial (Mullany et al., 2006).

in observational data to understand how CHX treatment effects depend on community- and individual characteristics — which vary much across our nationally representative sample — and then use this understanding to identify an optimal targeting policy. Meta-analyses of existing randomized trials have concluded that CHX cord care was only effective for home deliveries, which is not surprising given that 90% or more of the births included in the South Asian trials took place at home vs. between 36% and 47% in the Southeast African trials (Imdad et al., 2013; Sankar et al., 2016; López-Medina et al., 2019). But these meta-analyses have important limitations due to the small number of included studies and the possibility that heterogeneous results by place of birth may be confounded by other differences across studies — such as the number of CHX applications or factors correlated with local economic development or cultural practices. The use of ML methods to study heterogeneous treatment effects has many advantages. In particular, it allows a high degree of flexibility in identifying sources of heterogeneity and has embedded robustness checks in the form of cross-validation (Varian, 2014; Athey and Imbens, 2016).

Our third contribution is to take the ML findings obtained in our nationally-representative, Nepalese observational data, and use them to predict treatment effects in the five regions in- and outside Nepal where CHX trials have been carried out. In doing so, we replicate the broad pattern of heterogeneity observed in prior experimental findings across five different countries. Recent work studies heterogeneous treatment effect patterns to address questions of external validity in treatment evaluations in a variety of ways. Meager (2019) and Vivalt (2020) apply Bayesian hierachical models to experimental data from multiple RCTs. Bruhn (2020) uses a hierarchical Bayesian estimator combining experimental and non-experimental variation and finds that complier selection drives a large wedge between the Local Average Treatment Effect (LATE) observed in the experiment and the average treatment effect. Finally, Kowalski (forthcoming) estimates heterogeneous treatment effects from a randomized experiment (in Oregon) and uses them to predict the LATE obtained in a natural experiment (in Massachusetts) where compliers are different. But to the best of our knowledge this paper is the first to leverage heterogeneity patterns identified in a natural experiment to predict the heterogeneous findings reported across different RCTs — thus circumventing issues of RCT microdata availability and treatment comparability (e.g., number, timing and dosage of CHX applications)

across different RCTs.

The first country to introduce CHX cord cleansing nationwide is Nepal. We exploit plausibly exogenous variation in the timing of the CHX cord care program expansion across districts of Nepal using data from the nationally representative 2016 Nepal Demographic and Health Survey (DHS). After piloting the program in 4 out of 75 districts from late 2009, CHX cord application was quickly scaled-up across the rest of the country (see Figure 1). By 2015, 75 percent of the population was covered by the program (Department of Health Services, 2015). While the Chlorhexidine Navi(Cord) Care Program was integrated to the training, monitoring systems and operations of a recent national newborn health program ("Community-Based Newborn Care Program" or CB-NCP), the timing at which CHX was rolled out to a particular district largely depended on practical considerations such as presence of implementation partners on the ground and district government leadership.[4]

Twoway fixed effects estimates indicate that, overall, the CHX program decreased neonatal mortality by 1.8 percentage points or 43 percent compared to the control group mean. This was driven by a 56 percent decrease among births predicted to take place at home, while the estimated effect is both very small in magnitude and statistically insignificant among babies predicted to have been delivered in health facilities. Our conclusions are robust to a comprehensive number of robustness checks.

Our results show that children only benefit from CHX application, on average, if they are predicted to be born at home, in line with WHO recommended use. Place of delivery — which we correctly predict in 76 percent of cases for which we observe actual place of delivery — is however likely to proxy for risk factors such as hygiene conditions and healthcare at- and shortly after birth, health endowment at birth, and treatment compliance. To better describe the treatment effect heterogeneity we observe, we turn to machine learning. Our causal forest detects significant heterogeneity in treatment effects, and when comparing the lowest- with the highest treatment effect quartiles, we find large, statistically significant treatment effects in the two top quartiles but no significant effect in the rest of the sample. Looking closer at the characteristics of the samples in the top- and bottom quartiles of the distribution of treatment effects,

[4]Implementation partners were Care Nepal, Save the Children, Health Right International, UNICEF, ADRA and One Heart Worldwide (JSI Research & Training Institute, 2017).

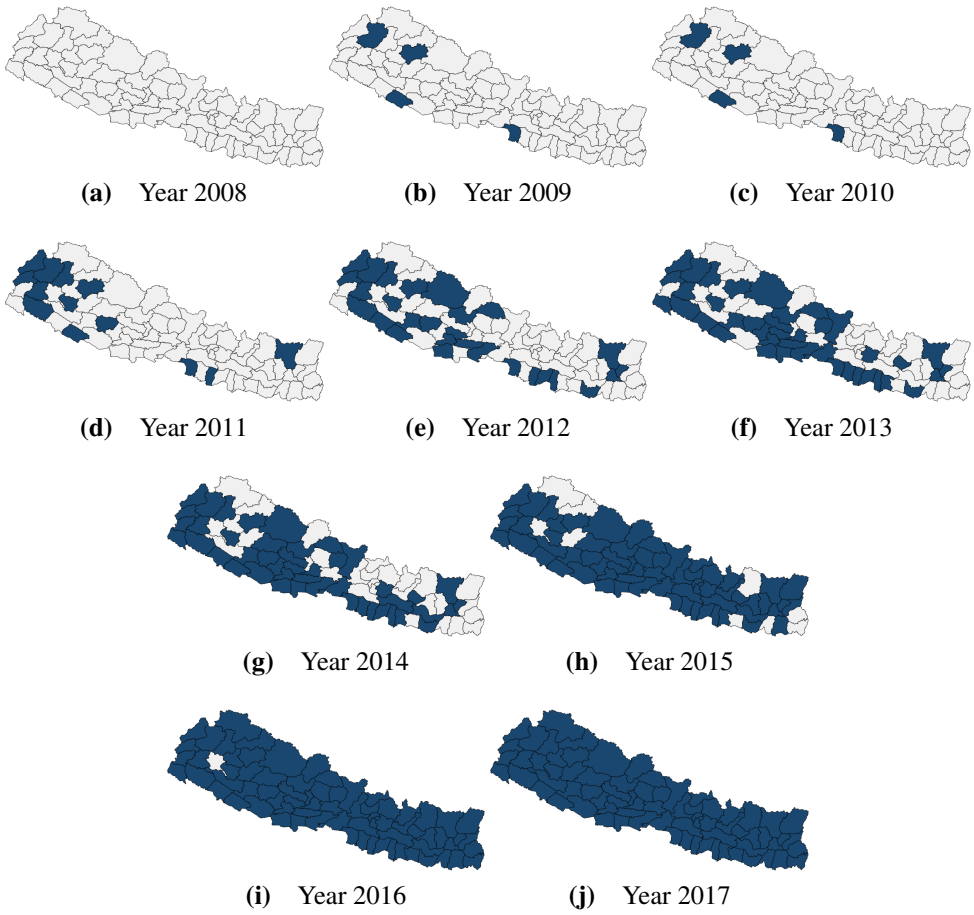


Figure 1: CHX cord application roll-out across districts over time (adopted CHX=blue).

we find that children in the highest treatment effects quartile are more often boys (unsurprisingly given that male newborns are more likely to die), and are often born to very young, less educated, rural mothers, and more often born at home.

We then apply Athey and Wager (2021)'s approach to identify the targeting policy which would result in the largest gains in neonatal survival, and compare this optimal policy — from the point of view of neonatal survival — to the WHO recommendations. We find that the WHO recommended policy of targeting all births taking place at home in areas (here, districts) with a neonatal mortality of 30 per thousand and above near-optimally identifies the third or so of the population with the highest returns to treatment. However, it misses many other children whose survival chances could significantly benefit from CHX treatment and therefore only achieves half of the overall reduction in neonatal mortality

achieved by a policy optimally targeting districts based on three other district-level indicators of health care quality.

Finally, after applying the causal forest to our nationally representative Nepalese dataset, we take advantage of the international comparability of the DHS and predict average treatment effects (ATEs) based on observable characteristics in five different DHS samples corresponding to the subnational regions and time periods where the five CHX RCTs took place. Doing so, we predict a similar pattern of ATEs to what is obtained experimentally in the five RCTs, with ATEs between -2.6 and -1.7 percentage points in the three South-Asian settings where large reductions in NMR were observed among children treated with CHX and ATEs between only -0.6 and -0.5 in the two Southeast African settings where no significant effect was found.

Prior work studying the effect of health programs carried out at scale in developing countries has found at best small decreases (<0.3 percentage points) in neonatal mortality (Lim et al., 2010; McKinnon et al., 2015; Powell-Jackson et al., 2015; Arulampalam et al., 2017; Van de Poel et al., 2016; Philibert et al., 2017; Fitzpatrick, 2018). Broad-based National Health Insurance systems introduced in the last few decades in a number of middle-income countries have been found to reduce infant mortality (see Conti and Ginja, forthcoming, and references therein), but the few estimates on neonatal mortality are mixed (Bhalotra et al., 2019).[5,6] In this context, CHX appears to offer a valuable option to health policy makers looking for evidence-based, affordable, at-scale solutions to reduce neonatal mortality in low-income settings.

In the next section, we give an overview of early life mortality trends and CHX cord care in Nepal. Section 3 presents the data and identification strategy. The main regression results and robustness checks are reported in Section 4. Section 5 explores heterogeneity in the effect of CHX application using machine learning, draws lessons for targeting, and extrapolates our quasi-experimental heterogeneity analysis to reconcile reported findings across RCT studies. Section 6 concludes.

[5]PROGRESA, which paid cash transfers conditional to poor households conditional on, among others, regular prenatal checks, has been found to significantly decrease infant mortality, but not neonatal mortality (Barham, 2011).

[6]Historical evidence from today's developed countries has concentrated on infant mortality (see, e.g., Bauernschuster et al., 2017, and references therein).

2 Background

2.1 Early Life Mortality in Nepal

Nepal is a landlocked country situated between China and India which is home to 28.1 Million people. The country's Human Development Index ranks only 147 out of 189 (in 2019) and more than a third (36 percent) of children under age 5 are stunted. Notably, Nepal has seen sharp decreases in fertility over the past twenty years — from 4.6 children per woman in 1995 to 2.3 in 2016, and marked reductions in child mortality — from 118 deaths before the age of 5 per 1,000 births in 1992-1996 to 39 in 2012-2016.

However, progress in the NMR reduction in Nepal stalled in the early 2000s (at 33 per 1,000 both during 2002-2006 and 2007-2011) while under-5 mortality slowed down its downward trend, going from 61 to 54 per 1,000 during the same period.[7] This stagnation came to an end in 2012-2016 as NMR dropped to 21 per 1,000 — a 36 percent decline relative to the previous 10-year period.

2.2 Chlorhexidine Cord Care

The latest decrease in NMR observed since 2012 coincides with the acceleration of the roll-out of CHX cord application through the Chlorhexidine Navi(Cord) Care Program (CHX-NCP) (see Figure 2).

CHX-NCP was a $3.9 million program funded mainly by bilateral donors (US, Norway, Canada, UK) and the Bill & Melinda Gates Foundation. In partnership with the Nepalese Department of Health Services, international NGOs and a Nepalese pharmaceutical company which produced the CHX gel locally, the program was implemented by JSI Research & Training Institute, Inc. and was designed to support the Government of Nepal to scale up the use of CHX for cord care nationwide. This involved training as well as procurement, logistical, monitoring and technical support.

The scaled-up intervention consisted of a single CHX gel application on the day of birth to all newborns irrespective of place of birth. For home births, CHX gel doses were distributed to pregnant women during antenatal care visits

[7] All the mortality and fertility figures in this sub-section are taken from Ministry of Health [Nepal] and New ERA and ICF (2017a).

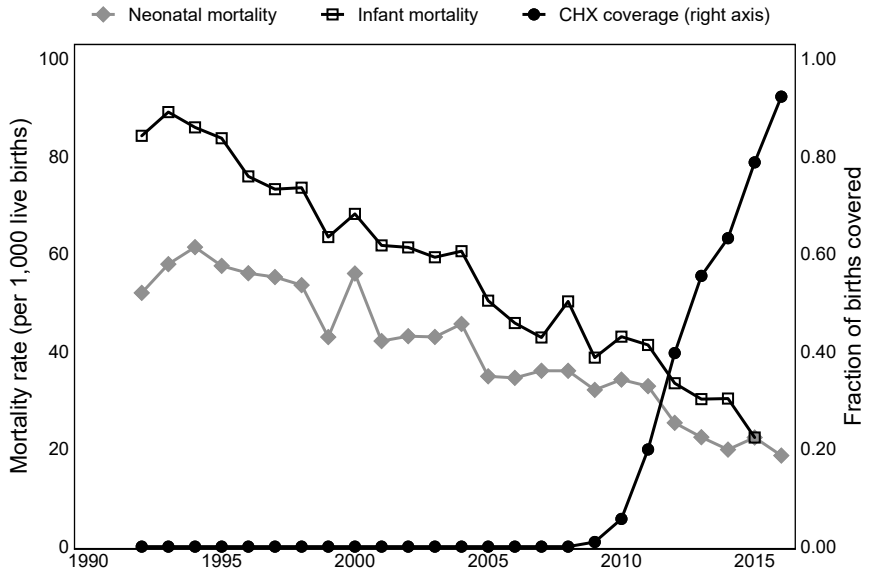


Figure 2: Child mortality and CHX-NCP coverage
Source: Own calculation based on DHS 2016 merged to administrative records on the roll-out of CHX-NCP.

in the last two months of pregnancy (Hodgins et al., 2019).[8,9] The CHX training of health workers lasted between three hours and one day and to reduce costs and increase program sustainability, training and monitoring activities were integrated into broader maternal and newborn health programs, and in particular into the Community-Based Newborn Care Program (CB-NCP) (JSI Research & Training Institute, 2017; JSI, 2017; Hodgins et al., 2019).

Estimates of actual CHX application in program districts vary much and, for home deliveries, an important limitation is that there is no record of application and that maternal recall is unlikely to be reliable for non-salient events (Beckett et al., 2001).[10] Coverage estimates suggest that it may have peaked in

[8]Eighty four percent of women who gave birth in the five years leading to the 2016 DHS received antenatal care and 69 percent received four antenatal care visits or more (Ministry of Health [Nepal] and New ERA and ICF, 2017a).

[9]Table A.1, Panels B and C report results obtained when estimating our twoway fixed effects equation using, in turn, the number of antenatal visits or an indicator for having an above-median number of antenatal visits as dependent variable, which show that CHX-NCP was not accompanied by an increase (or decrease) in the number of antenatal care visits.

[10]In the DHS, women who gave birth within five years of the interview are asked, among many other things, whether anything was placed on the stump after the umbilical cord was cut, and if so, what substance was applied. There is good reason to think that answers to these questions are not reliable: While CHX was neither available nor promoted in a district prior to

2014/2015, as estimates range from 75 percent of home deliveries and 96 percent of facility deliveries (HIMS (2014), as cited in Khanal (2015)) to 75 percent of all births according to Department of Health Services (2015) to only about 40 percent of home births and 90 percent of facility births in 2017 according to Hodgins et al. (2019) so that estimates presented in this study should be interpreted as intention-to-treat effects — arguably the parameter of interest from a policy point of view. The coverage is however consistently estimated to be higher among health facility deliveries, so that heterogeneity in treatment intensity cannot account for the larger decrease in NMR observed among predicted home births.

3 Data and Identification Strategy

3.1 Data

The 2016 Demographic and Health Survey (DHS) of Nepal is a nationally representative survey that collected detailed pregnancy histories of all women age 15-49 found in sampled households, as well as comprehensive data on the demographic and socioeconomic characteristics of the household and its members. The dataset includes, for each child ever born to the interviewed women, dates (month and year) of birth and death, if applicable. Detailed information on antenatal and postnatal care is also collected for births occurring within 5 years of the interview, including place of delivery (Ministry of Health [Nepal] and New ERA and ICF, 2017b). In the absence of comprehensive vital statistics systems, the DHS is the main source of information on child mortality in Nepal as in many other developing countries.

The survey collected data on a total of 26,028 births. We drop 366 multiple births, 118 births to mothers who are either less than 15 or 45 and above and 118 births occurring within one month of the interview date and thus not fully exposed to the risk of neonatal death. While recall error is unlikely to be an issue for such a salient event in the life of a woman as the death of a newborn, we restrict our main analytical sample to births that occurred within 25 years

the roll-out of CHX-NCP, as many as 16 percent report that CHX was applied to the stump of the newborn in *untreated* district-by-time cells. Meanwhile only 31 percent report that CHX was applied to the stump of the newborn in *treated* district-by-time cells, which is less than half what is found in administrative records.

prior to the date of interview, resulting in a sample of 23,465 births. Robustness checks varying this time window by 5 years on either side show that our findings are not sensitive to this sample selection criteria (see Section 4.2).

We merge the DHS microdata with administrative data on the implementation of all the main programs targeting maternal and newborn health in Nepal which were not available in all districts of Nepal by 2009 when CHX-NCP was first piloted. It is to be expected that in any non-experimental setting, a number of initiatives from national authorities and international organizations are ongoing at any one time. To ensure that we captured the effect of CHX-NCP independently of any other intervention, a thorough identification of programs that may have contributed to recent decreases in NMR was done by the Kathmandu-based Center for Research on Environment, Health and Population Activities (CREHPA) in two steps. First, all annual reports produced by the Department of Health since 2013 were analyzed in detail to identify candidate explanations for the recent decrease in NMR. Second, semi-structured interviews with 12 in-country neonatal and maternal health experts — from, among others, the Family Welfare Division of the Department of Health Services, the WHO, UNICEF, and Children and Maternity hospitals — were carried out in order to collect their specialist views on the most likely reason(s) for the NMR reduction.[11] Dates of the district-level roll-out of each program were then collected from various Department of Health Annual Reports, and controls included in the main analysis for the two health programs targeting newborns specifically (CB-NCP and CB-IMNCI) and in robustness checks for secondary programs whose coverage is not fully captured by time fixed-effects. For CHX-NCP, which was administered by JSI, we obtained roll-out dates from the CHX-NCP program director.

In Table 1, we report summary statistics for the whole sample and separately for children predicted to be born at home or not to be born at home using the approach described in Section 3.2. These statistics highlight that the sample at hand has very low levels of human development, with 57 percent of children having mothers with no formal education, 41 percent living in rural areas, and one in five children being born to a teenage mother. Forty-eight percent of children are female, which is close to what would be expected given the widely observed natural sex ratio at birth (51 percent male).

[11] Ten interventions were identified by key informants, including CHX-NCP.

Table 1: Variable means

		P(Home Birth)	
	All	<0.5	> 0.5
A. Demographics and SES			
Female	0.48	0.47	0.50
First born	0.34	0.53	0.14
Second born	0.28	0.29	0.27
Third born	0.18	0.11	0.25
Parity four or higher	0.21	0.07	0.34
Mother age 15-19y	0.20	0.26	0.13
Mother age 20-24y	0.41	0.46	0.37
Mother age 25-29y	0.26	0.21	0.32
Mother age 30-34y	0.10	0.07	0.13
Mother age 35-39y	0.03	0.01	0.04
Mother age 40-45y	0.01	0.00	0.01
Ethnicity Dalit	0.15	0.12	0.19
Rural	0.41	0.22	0.60
Education: no education	0.57	0.33	0.81
Education: primary	0.18	0.22	0.14
Education: secondary	0.19	0.34	0.04
Education: higher	0.06	0.11	0.01
Wealth in 1st quintile	0.27	0.07	0.48
Wealth in 2nd quintile	0.22	0.16	0.28
Wealth in 3rd quintile	0.20	0.25	0.15
Wealth in 4th quintile	0.17	0.27	0.07
Wealth in 5th quintile	0.13	0.25	0.01
B. Health programs			
Program: CB-NCP	0.16	0.18	0.13
Program: CB-IMNCI	0.05	0.07	0.03
Program: CHX	0.13	0.15	0.11
C. Child mortality			
Child died ≤1m	0.04	0.03	0.05
Child died <1m	0.03	0.03	0.04
Child died ≤12m	0.06	0.04	0.07
Child died ≤12m & >1m	0.01	0.01	0.02
Observations	23,465	11,719	11,746

Notes: Column two shows means for variables based on the full analysis sample. In columns three and four we split the sample according to the predicted place of delivery, based on the linear probability model shown in Appendix Table A.2.

3.2 Identification Strategy

In our main specification, we estimate linear probability models of the form:

$$m_{idt} = \alpha + \beta CHX_{dt} + D_d'\Delta + T_t'\Gamma + X_{idt}'\Lambda + \varepsilon_{idt} \tag{1}$$

where m_{idt} is an indicator equal to 1 if child i dies by age one month (allowing for "heaping" at one month) and zero otherwise, CHX_{dt} is an indicator equal to 1 if CHX-NCP was rolled out in the child's district by the date the child was born, D_d is a vector of district fixed effects, T_t is a vector of time fixed effects, where time is defined at the month-by-year level (e.g., Ashwin 2066 in the Nepali calendar or October 2009), X_{idt} is a vector of controls comprising child, mother, household characteristics and district-time varying controls such as exposure to health programs other than CHX-NCP; $\alpha, \beta, \Delta, \Gamma$ and Λ are parameters to be estimated; and ε_{idt} is an error term allowing for arbitrary intra-district correlation.

The WHO guidelines recommend the application of CHX to the cord only for home births in settings with neonatal mortality above 30 per 1000. Guided by the distinction made by the WHO between home- and facility deliveries, we allow for heterogeneous treatment effects across predicted home deliveries and predicted facility deliveries.

Place of delivery is only collected by the DHS for births in the five years leading to the survey. To use data covering a longer period of time and thus increase statistical power, we predict whether a child was delivered at home using a linear probability model regressing an indicator for being delivered at home on birth order, maternal age group, child gender, maternal ethnicity, altitude quintile, maternal education, rural location, wealth quintile, district fixed effects and date of birth — defined by Nepali month and year of birth — fixed effects (see Table A.2).

In the sample for which we know the place of delivery, when predicting a home birth based on a probability of home delivery above 0.5 predicts place of birth correctly in 76 percent of cases (see Appendix Figure A.1). In order to account for the uncertainty in classifying births based on their predicted- rather than observed place of delivery, we obtain bootstrapped standard errors — clustered at the district level — by drawing 200 random samples from the original dataset, and, for each random sample, predicting whether the baby is delivered at home or not and then re-estimating the relevant variant of Equation

(1).

Since we control for time- and district fixed effects, identification relies on the absence of time-varying omitted factors correlated with the timing of treatment. Regressing the treatment indicator on observable characteristics, we find that, other than the expected positive correlation between CHX-NCP and CB-NCP, the program on which CHX-NCP "piggy-backed" (Hodgins et al., 2019), the treatment is only weakly correlated with observable characteristics.[12] Among the sample of births predicted to take place in an institution, treated babies are significantly less likely to be found in rural areas, to be their mother's third born and more likely to have a mother with an ethnicity from the residual "other" group. Among the sample of births predicted to take place at home, babies born after CHX was introduced in their district are slightly — up to 1.6 percentage points — less likely to be born to a mother with a secondary degree and more likely to have a mother from the second wealth quintile. However, these differences are small, there is no clear pattern of selection in terms of socio-economic status and, in the case of predicted home births, only statistically significant at the 10 percent level (See Figure 3 and Appendix Table A.3). In Section 4.2, we report on a number of robustness checks which indicate that our findings are unlikely to be biased by a correlation between district trends in early life health and the timing of CHX-NCP roll out.

Recent work has shown that, in the presence of heterogeneous treatment effects, two-way fixed effects models such as the one we estimate can significantly depart from the average treatment effect (e.g., Goodman-Bacon, 2021; de Chaisemartin and d'Haultfoeuille, 2020). Checks reported in Section 4.2 suggest that our results are not driven by weighting issues in the two-way fixed effects model.

[12] When interacting our CHX program indicator with the CB-NCP indicator, the interaction effect is statistically insignificant, suggesting that complementarities between the effects of the two programs are limited.

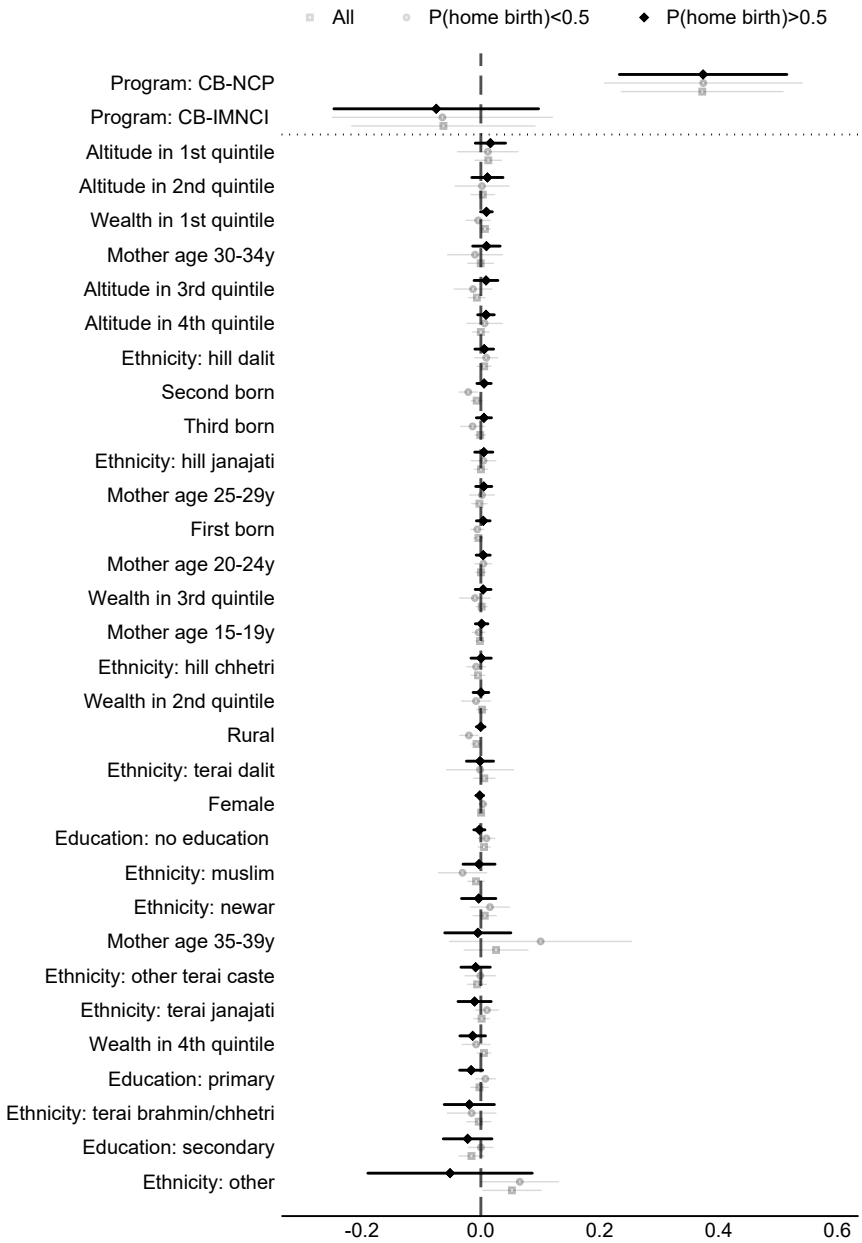


Figure 3: Covariate balance

Notes: This chart shows the beta coefficients and the 95 percent CI from running a regression with the treatment indicator as the dependent variable and each of the covariates listed in the figure, in turn as independent variables, as well as district and month of birth fixed effects. The confidence intervals are calculated based on standard errors obtained through 200 bootstrap iterations clustered at the district level. We split the sample according to the predicted place of delivery, based on the linear probability model shown in Appendix Table A.2. Appendix Table A.3 reports all coefficients.

4 Results

4.1 Main Results

Table 2 reports our baseline estimates. In Column (1), we estimate Equation 1 on the full sample and find that CHX-NCP decreases neonatal mortality by 1.8 percentage points or 43 percent of the control mean. This is larger than the 20-34% decreases observed in the three Southeast Asian RCTs, which suggests that the additional preventive and remedial measures in place as part of the RCTs may have contributed to limit the benefits of CHX application. In Column (2), we allow the effect of CHX-NCP to vary by predicted place of birth by including a control for predicted place of birth (1[P(home birth)>0.5]) and an interaction between predicted home birth and the CHX-NCP treatment variable. In this specification, the treatment effect is not significant for predicted facility deliveries but it is four times larger (2.8 percentage point) and statistically significant among predicted home deliveries. Finally, in Columns (3) and (4) we allow all the model coefficients to vary by predicted place of birth, which leads to a near-zero estimated effect of CHX-NCP among predicted facility deliveries (0.1 percentage point) while the estimated decrease in the probability of neonatal mortality among predicted home deliveries remains equal to 2.8 percentage points — and we can reject the null of no difference in treatment effect between the two samples defined by predicted place of delivery (p-value: 0.031).

CHX-NCP covered both home- and facility deliveries and therefore it did not create an incentive for mothers to deliver at home rather than in a facility or vice-versa in order to obtain a CHX dose. For the subsample for which we know the place of birth, we can test whether CHX-NCP had an effect on place of birth. Table A.1, Panel A reports results obtained when estimating Equation (1) using an indicator for home delivery as dependent variable, which show that CHX-NCP did not change the probability of a home delivery.

4.2 Robustness Checks

We start by addressing the question of whether our treatment effect captures unobserved time-varying factors associated with a decrease in NMR in treated districts relative to control districts. To do so, we carry out three checks which bolster our confidence in the causal interpretation of our results.

Table 2: Regression results: The effect of CHX-NCP on neonatal mortality - Dependent variable: Mortality by ≤1m.

	Sample			
	All	All	P(home birth) <0.5	P(home birth) >0.5
	(1)	(2)	(3)	(4)
CHX	-0.018**	-0.007	0.001	-0.028**
	(0.007)	(0.007)	(0.009)	(0.011)
1[P(home birth)>0.5]		-0.001		
		(0.005)		
CHX × 1[P(home birth)>0.5]		-0.021***		
		(0.008)		
CHX + CHX × 1[P(home birth)>0.5]		-0.028***		
		(0.008)		
Observations	23,465	23,465	10,860	12,605
Clusters	73	73	73	73
Control mean of dep. var	0.042	0.042	0.033	0.050
P-val (dif across sample)				0.031

Notes: All specifications are estimated as linear probability models using OLS with the full set of demographic, SES, and program controls. Demographic controls include birth order (three indicators), five year maternal age group indicators, and gender. SES controls include education (three indicators), wealth (four indicators), rural indicator, altitude quintile indicators, and ethnicity indicators. Program controls include controls for the CB-NCP and CB-IMNCI health programs. All coefficients are reported in Appendix Table A.4. All specifications are estimated with district and month of birth fixed effects. We split the sample according to the predicted place of delivery, based on the linear probability model shown in Appendix Table A.2. Bootstrapped standard errors based on 200 iterations and clustered at the district level in parentheses. Asterisks indicate significance at the following levels: * p<0.1, ** p<0.05, and *** p<0.01.

First we define a placebo treatment which is equal to one if the child was born 12 months before the CHX-NCP was rolled out in the district or later, and zero if the index child was born earlier. In Column (1) of Table 3, we show that including this variable in our main specification has no effect on our estimated treatment effect, and that the coefficient associated with the placebo treatment variable is very close to zero (0.001) and statistically insignificant.[13]

Second, we re-estimate Equation 1 using mother fixed-effects instead of dis-

[13]If instead we keep only pre-treatment observations and estimate Equation 1 replacing CHX_{dt} with the placebo treatment CHX_{dt-12}, we also find that the placebo treatment effect is close to zero (0.003) and statistically insignificant (Table A.5), whereas if we estimate Equation 1 in a sample including only untreated children and children born no more than 12 months after the roll-out of CHX-NCP in their district, the effect of CHX-NCP is close to our main estimate (-0.027) and statistically significant at the 5 percent level (Table A.6).

Table 3: Regression results: placebo treatment, mother fixed effects, infant mortality. Sample: P(home birth)>0.5

	Placebo	Mother FE	Mortality	
			>1m&≤12m	≤12m
	(1)	(2)	(3)	(4)
CHX	-0.029**	-0.032**	0.015**	-0.016
	(0.012)	(0.016)	(0.006)	(0.014)
CHX_{t-12}	0.001			
	(0.013)			
Observations	12,605	11,654	12,373	12,373
Clusters	73	73	73	73
Control mean of dep. var	0.050	0.050	0.018	0.068

Notes: All specifications are estimated as linear probability models using OLS with the full set of demographic, SES, and program controls on the sample with P(home birth)>0.5. Demographic controls include birth order (three indicators), five year maternal age group indicators, and gender. SES controls include education (three indicators), wealth (four indicators), rural indicator, altitude quintile indicators, and ethnicity indicators. Program controls include controls for the CB-NCP and CB-IMNCI health programs. All specifications are estimated with district and month of birth fixed effects. The place of delivery is predicted using the linear probability shown in Appendix Table A.2. Bootstrapped standard errors based on 200 iterations and clustered at the district level in parentheses. Asterisks indicate significance at the following levels * p<0.1, ** p<0.05, and *** p<0.01.

trict fixed effects and find similar results (Table 3 Column (2)). This indicates that our district fixed-effects estimates are not biased by differential changes in the composition of mothers between treated and control districts (e.g., due to differential trends in maternal education or living standards between maternal cohorts).

Third, we carry out a falsification test based on the fact that cord infection (omphalitis) primarily affects neonates, but is uncommon among older infants (Painter and Feldman, 2019). CHX application, which narrowly targets omphalitis, should therefore decrease neonatal mortality but not mortality between 2 and 12 months of age — whereas unobserved time-varying improvements in maternal and child health should decrease both. In Column (3) of Table 3, we estimate Equation 1 using as dependent variable an indicator equal to 1 if the child died between 2 and 12 months of age and zero if they survived beyond infancy — the 12 first months of life — and find that babies born under the CHX-NCP program were *more* likely to die between 2 and 12 months. This is both interesting and unsurprising: risk factors for the development of omphalitis

include a number of risk factors for post-neonatal infant mortality such as low birth weight and unhygienic practices (Painter and Feldman, 2019) so that the babies who survive the neonatal period due to CHX are "negatively selected" — i.e., disproportionately likely to die later in infancy. Reassuringly, the total effect of CHX-NCP on overall mortality in the first year of life is however a statistically insignificant but large in magnitude *decrease* in infant mortality (by 1.6 percentage points).

Given the small sample sizes we have in our data at the monthly level — the level at which treatment is defined, an event-study analysis leads to very imprecise estimates. For completeness, we report the estimates obtained from an event-study analysis at the quarterly level (Figure 4), which show a noisy but largely flat and non-negative pattern prior to the introduction of the CHX program in the district, and then increasingly negative treatment effects after the program is rolled out. Similarly, we lack statistical power if restricting the sample to recent births, for which we know the place of delivery — especially when splitting the sample by place of birth rather than using the whole sample and interacting place of birth only with CHX-NCP. Results are however qualitatively similar. First, the treatment effect obtained on this sample, while statistically insignificant, corresponds to a reduction of 29- to 39 percent of the control mean among home births (Columns (4) and (2) respectively, Table A.7). Second, no reduction in the probability of neonatal mortality is observed for babies delivered in a facility and third, the difference between the treatment effects for home- and facility births is statistically significant.

Recent work has shown that, in the presence of heterogeneous treatment effects, two-way fixed effects models such as the one we estimate can significantly depart from the average treatment effect (e.g., Goodman-Bacon, 2021; de Chaisemartin and d'Haultfoeuille, 2020).[14] Of particular concern is the fact that some of the treatment effects averaged over in the two-way fixed effects model bear negative weights. To address this issue we compute the weights derived in de Chaisemartin and d'Haultfoeuille (2020) and find that 16 percent of our 862 weights are negative, and have a total weight of -0.054 (Table 4).

Reassuringly, we compute the minimum standard deviation in the treatment

[14]Note that the concern about treatment effect heterogeneity raised by Goodman-Bacon (2021) and de Chaisemartin and d'Haultfoeuille (2020) applies to treatment effect heterogeneity between cells defined here by district and month/year, not between individuals *within* district-time cells differing, e.g., by predicted place of birth or other individual characteristics.

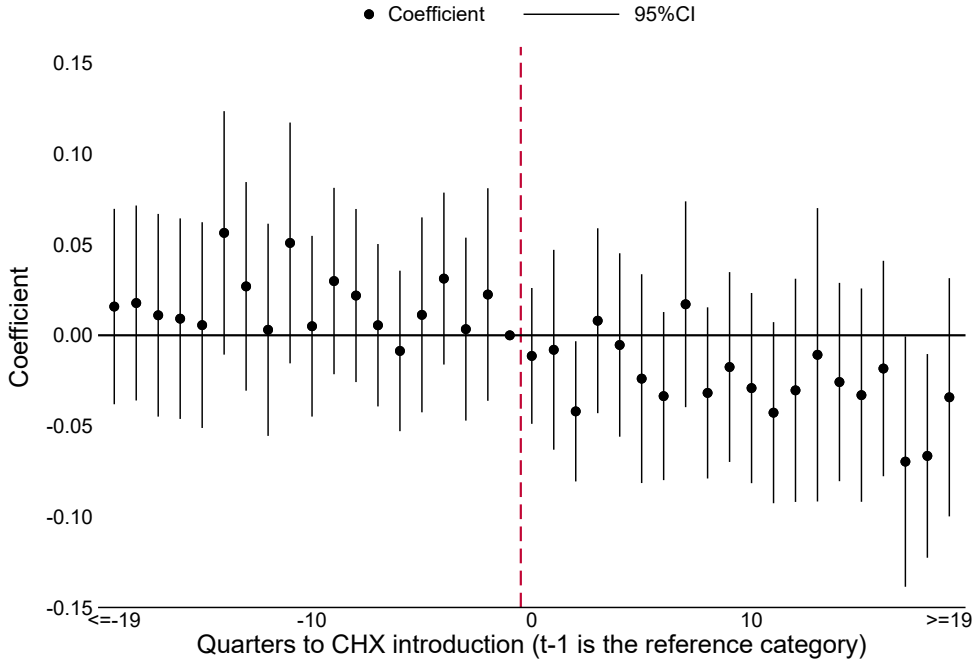


Figure 4: Event study chart.
Notes: Estimated on the sample: P(home birth)>0.5, with the full set of demographic, SES and program controls, as well as month of birth and district fixed effects. The confidence intervals are calculated based on standard errors obtained through 200 bootstrap iterations clustered at the district level. The place of delivery is predicted using the linear probability shown in Appendix Table A.2.

effect across all district-month cells which would be required for the average ATT over all cells to in fact be zero, and find that the required amount of heterogeneity is implausibly large.[15] As shown in Table 4, the minimum standard deviation required is 0.0307 (Column 5). The pre-treatment NMR incidence in the predicted home-birth sample is 5 percent or 0.05. If the (absolute values of the) ATTs for our 862 cells were drawn from a uniform distribution between 0 (no effect) and 0.05 (total eradication of NMR), the standard deviation (SD) would only be 0.014. If they were drawn instead from a normal distribution with mean 0.028 (our two-way fixed-effects estimate) and SD 0.0307 — the minimum SD required for the average ATT over all cells to be zero, we would have 40 percent of ATTs to be outside the [0,0.05] range, which is not plausible. In addition, after dropping the 140 cells with negative weights, the estimated

[15] We compute both the weights and the minimum standard deviation using de Chaisemartin and d'Haultfoeuille (2020)'s *twowayfeweights* command.

effect of CHX application is almost identical (-0.029). In this new sample, the weights change and 7 percent of cells now have negative weights. After five iterations of dropping cells with negative weights and re-estimating both our two-way fixed effects model and the remaining cells weights, we obtain a sample with no negative weights and the treatment effect on the remaining cells is -0.026, compared to -0.028 in the full sample, demonstrating that our results are not driven by the negative weighting of some treatment effects.

We also estimated a number of alternative specifications for Equation (1) and found no notable difference in estimates. In these alternative specifications, we removed all controls other than district and time effects, varied the subsets of controls included, added controls for additional health and nutritional programs, *in-utero* exposure to the severe earthquake which took place in 2015, controlled for an interaction term between baseline district neonatal mortality and a linear

Table 4: Assessing the role of negative weights in the two-way fixed effects estimator.

	Coefficient (1)	N_w (2)	$N_{w<0}$ (3)	$\sum_{w<0} w$ (4)	σ_{FE} (5)
Baseline	-0.028**	862	140	-0.053974	0.0307
	(0.011)				
Iteration 1	-0.029***	662	46	-0.008862	0.0394
	(0.010)				
Iteration 2	-0.028***	600	12	-0.000816	0.0419
	(0.010)				
Iteration 3	-0.027**	582	8	-0.000241	0.0425
	(0.011)				
Iteration 4	-0.026**	570	4	-0.000034	0.0415
	(0.011)				
Iteration 5	-0.026**	565	0	0.000000	NA
	(0.011)				

Notes: All models are estimated on the sample: P(home birth)>0.5, with the full set of demographic, SES, and program controls as well as district and month of birth fixed effects. Demographic controls include birth order (three indicators), five year maternal age group indicators, and gender. SES controls include education (three indicators), wealth (four indicators), rural indicator, altitude quintile indicators, and a Dalit ethnicity indicator. Program controls include controls for the CB-NCP and CB-IMNCI health programs. The place of delivery is predicted using the linear probability shown in Appendix Table A.2. Bootstrapped standard errors based on 200 iterations and clustered at the district level in parentheses. Asterisks indicate significance at the following levels * p<0.1, ** p<0.05, and *** p<0.01.

trend in month-year date of birth, and varied the sample in two ways: (i) changing the time period covered by the data — adding and removing five year cohorts on either side of our baseline 25-year panel — and (ii) removing or not children for whom the district of birth cannot be established with certainty because their mothers were currently visiting the household surveyed or because the woman had moved to the district where she was interviewed after the CHX program was first introduced in the country. As depicted in Figure 5, the estimated treatment effect for predicted home deliveries is consistently between -0.022 and -0.030 across specifications.[16]

Our findings are also robust to adopting an alternative definition of neonatal mortality which is equal to zero for children reported to have died at exactly one month old and which are counted as having died within the neonatal period in the main analysis to allow for heaping (Appendix Table A.8). Weighted least squares estimates using the sampling weights provided by the DHS also lead to the same conclusions (Appendix Table A.9).

Finally, we fitted a logistic model to reflect the binary nature of our dependent variable of interest. The estimated treatment effects are, again, similar to our main specification despite being larger in magnitude and less precisely estimated (Appendix Table A.10).

Our results show clear evidence of beneficial effects of CHX-NCP on children predicted to being born at home, and no evidence of such benefits, on average, among other births. Heterogeneous effects between predicted place of delivery may come from different rates of compliance or different efficacy conditional on compliance, for which place of delivery acts as proxy. Here we estimate ITT effects, which are of interest to policy makers when compliance cannot be enforced, as is the case for home births. As discussed in Section 2.2, compliance estimates vary but are close to 100% in the case of facility births, where treatment effect estimates are smallest, thus suggesting that different compliance rates are not a key driver of heterogeneity. In the next section, we investigate further the question of which babies should be targeted by CHX cord care programs to achieve the largest neonatal survival gains.

[16]Similarly, for children predicted to be born in a health facility, our estimates are consistently between -0.003 and 0.007 across the same specifications (see Appendix Figure A.2).

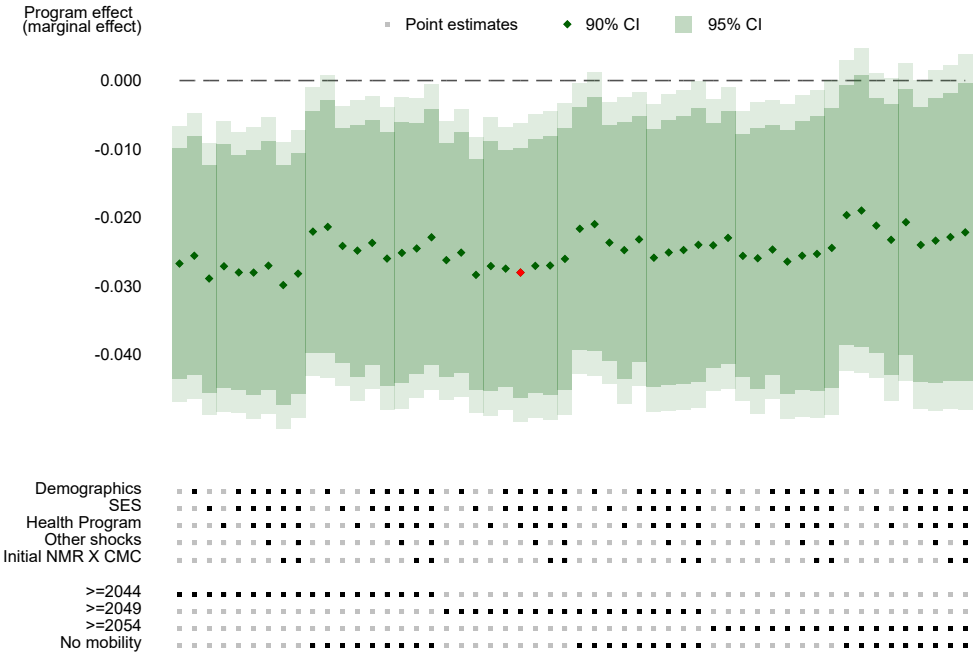


Figure 5: Specification Curve for predicted home births
Notes: This chart shows estimates from running 54 different specifications defined by the combination of markers bellow the chart. The red point indicates out main specification. Demographic controls include birth order (three indicators), five year maternal age group indicators, and gender. SES controls include education (three indicators), wealth (four indicators), rural indicator, altitude quintile indicators, and a Dalit ethnicity indicator. Program controls include controls for the CB-NCP and CB-IMNCI health programs. Other shocks refer to the earthquake on 25 April 2015, the Community Action for Nutrition Project, an Integrated Nutrition Program, and the Safe Delivery Incentive Program. Initial NMR $\times$ CMC is the initial neonatal mortality times a quadratic time trend. The place of delivery is predicted using the linear probability shown in Appendix Table A.2. The confidence intervals are based on 200 bootstrap iterations clustered at the district level.

5 Treatment Effect Heterogeneity and Lessons for Policy Targeting

5.1 Heterogeneity with Causal Forests

So far we studied treatment effect heterogeneity by place of delivery based on the WHO recommendation to treat children born at home (in districts with neonatal mortality rates above three percent). Place of delivery is however

likely to proxy for risk factors such as hygiene conditions and healthcare at- and shortly after birth and health endowment at birth (rather than differences in compliance, as just discussed). To better understand the treatment effect heterogeneity we observe and therefore potentially improve on current WHO recommendations for targeting, we turn to machine learning. We focus on characterizing heterogeneity along dimensions that can be realistically used by policy makers to target beneficiaries.

We first use regression forests to "residualize" the treatment indicator and our outcome of interest (neonatal mortality) — i.e., to purge them of variation coming from, in our case, availability of other health programs, and district- and month$\times$year of birth as captured by fixed effects. Using these residuals as outcomes, we then estimate a causal forest on a range of potential outcome predictors or "features". From a policy point of view, it may be more appealing to target whole communities than individuals, so in addition to the individual characteristics included as covariates in our two-way fixed fixed effects regressions, we also consider district-level features. Namely, we use, at the individual level: birth order, gender, maternal education, maternal age, wealth, altitude, rural, predicted place of delivery, health programs, district, ethnicity; and, at the district average level: antenatal care (ANC) visits (timing and number), whether iron tablets were received during ANC visits, tetanus protection, place of delivery, postnatal visits, immunization rate, neonatal mortality rate, delivery support by a nurse or doctor, and share of babies considered small at birth.

The building blocks of the causal forest are its trees. Each tree is created by partitioning a 50% draw of the sample into leaves defined by the value taken by a subset of features. The partitioning algorithm finds the combination of values taken by these features which maximizes treatment heterogeneity across leaves and penalizes treatment effect variance within leaves (Athey and Imbens, 2016). Following best practice, the fine-tuning of the algorithm is done optimally without researcher input based on cross-validation.[17]

Before reporting on the heterogeneity patterns uncovered by this exercise,

[17]We estimate the causal forest using the *grf* package in R (Tibshirani et al., 2021) with 2000 trees and all other parameter settings selected based on cross-validation. We use 50% of the sample to grow each tree. The splitting structure of the trees is determined on a 50% sub-sample of the tree sub-sample, after which the tree is populated by the the other 50% to estimate the treatment effects. For the splits in the trees we consider 30 variables and we restrict the nodes to have at least 5 observations. Appendix B provides further details about the causal forest procedure.

we report results of diagnostic tests which indicate that the causal forest successfully captures both average and heterogeneous treatment effects. More specifically, in panel A of Table 5 we show results of Chernozhukov et al. (2020)'s omnibus test for heterogeneity modified to be applied in an observational setting following the procedure implemented in Tibshirani et al. (2021). Intuitively we are estimating a linear regression of the individual's treatment effect predicted by the forest on the average predicted treatment effect (Mean Forest Prediction) and the individual's predicted deviation from the average treatment effect (Differential Forest Prediction). If the forest captures the average treatment effect well and if there is treatment effect heterogeneity that is also captured by the causal forest, both coefficients should be 1. In our case 1 is included in the confidence interval and there is evidence of significant treatment effect heterogeneity.

Table 5: Causal Forest Fit & Doubly Robust Average Treatment Effects

A. Omnibus diagnostic test for forest fit	
Mean Forest Prediction	1.215***
	(0.273)
Differential Forest Prediction	0.806*
	(0.525)
B. Doubly Robust Average Treatment Effects	
Full sample	-0.020***
	(0.003)
Predicted home delivery	-0.030***
	(0.004)
Predicted facility delivery	-0.007*
	(0.004)

Notes: Panel A shows the results for the omnibus test inspired by equation 3.1 in Chernozhukov et al. (2020) modified to the observational setting and implemented through the *test_calibration* function from the *grf* library in R. Panel B shows the Augmented Inverse-Propensity Weighted (AIPW) Average Treatment Effects. Standard errors in parentheses. Asterisks indicate significance at the following levels * $p<0.1$, ** $p<0.05$, and *** $p<0.01$. Note that the significance levels in panel A. are for the one sided tests.

Panel B of Table 5 shows the Augmented Inverse-Propensity Weighted (AIPW) Average Treatment Effects based on the causal forest. The AIPW is doubly robust, meaning that it is a consistent estimator of the ATE as long as at least one of (i) the propensity score *or* (ii) the outcome model, is correctly spec-

ified. Reassuringly, our AIPW estimates are similar to our two-way fixed fixed effects specification (full sample: -2 percentage points compared to 1.8 percentage points in Table 2, predicted home deliveries: -3 percentage points compared to 2.8 percentage points in Table 2), even though for predicted facility births, the AIPW is suggestive of CHX being somewhat effective (-0.7 percentage points, significant at 10% compared to an insignificant 0.1 percentage points in Table 2).

In Figure 6 we look at heterogeneity patterns by subgroups defined by their individual conditional average treatment effects (CATEs) — i.e., the AIPW predicted for individuals with the same characteristics. For the first and second quartiles, the estimated average treatment effect are large and statistically significant (-0.042 and -0.022 for the first and second quartile, respectively). But there is no detectable average treatment effect for the third and fourth quartiles. Moreover we reject the null-hypothesis of equal treatment effects across the four quartiles (p<0.001).

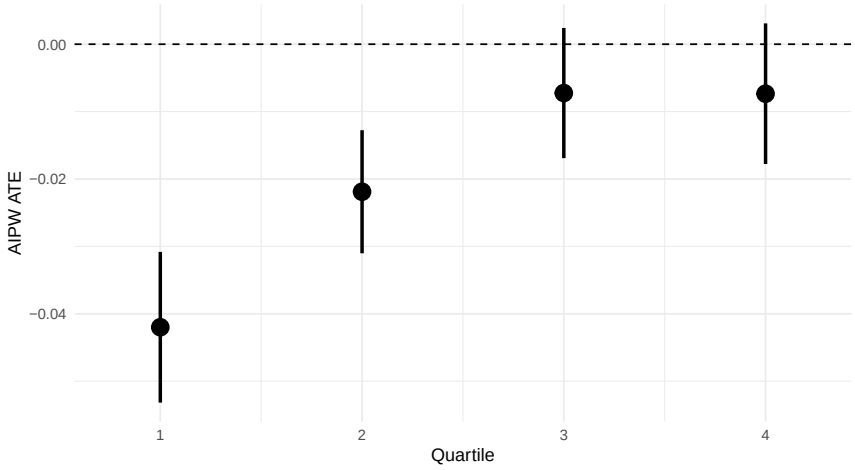


Figure 6: Doubly Robust ATEs for quartiles of CATEs

Notes: The Augmented Inverse-Propensity Weighted (AIPW) Average Treatment Effects are estimated for quartiles of the conditional average treatment effects shown in Appendix Figure B.5. P-val for H_0 of equal treatment effects: p<0.001.

As both the omnibus test results shown in Table 5 and the comparison of average treatment effects in Figure 6 show that there is evidence of substantial treatment effect heterogeneity, we now turn to a characterization of this heterogeneity. In Table 6 we compare means for selected variables across the first

and fourth quartiles. In the first quartile, where we observe large benefits of the treatment, 87 percent of births are predicted to take place at home compared to only 12 percent in the fourth quartile. Moreover, children in the first quartile are more often boys (unsurprisingly given that male newborns are more likely to die), and are often born to very young, less educated, rural mothers. Interestingly, the differences across quartiles do not simply reflect differences in baseline neonatal mortality rates: the CATE goes from -0.022 to -0.007 percentage points when going from the second to the third quartile although baseline NMR only decreases from 0.036 to 0.029.

Table 6: Covariate means across quartiles of CATEs

	Quartile			
	First	Fourth	Difference	P-val
Female	0.387	0.562	0.175	<0.001
Home delivery	0.707	0.240	-0.467	<0.001
Predicted home delivery	0.874	0.117	-0.757	<0.001
Age: 15-19y	0.283	0.088	-0.195	<0.001
Age: 20-24y	0.394	0.491	0.097	<0.001
Age: 25-29y	0.203	0.301	0.098	<0.001
Age: 30-34y	0.088	0.096	0.008	0.117
Age: 35-39y	0.027	0.021	-0.006	0.036
Age: 40-45y	0.005	0.002	-0.003	0.011
Education: No education	0.915	0.131	-0.784	<0.001
Education: Primary	0.081	0.148	0.067	<0.001
Education: Secondary	0.004	0.542	0.538	<0.001
Education: Higher	0.001	0.179	0.179	<0.001
Rural	0.596	0.229	-0.367	<0.001
Baseline NMR	0.049	0.028	-0.022	<0.001

Notes: The table shows covariate means for the first and fourth quartile of the sample based on the estimated CATEs. Baseline NMR is the district level neonatal mortality rate in the five years before treatment started.

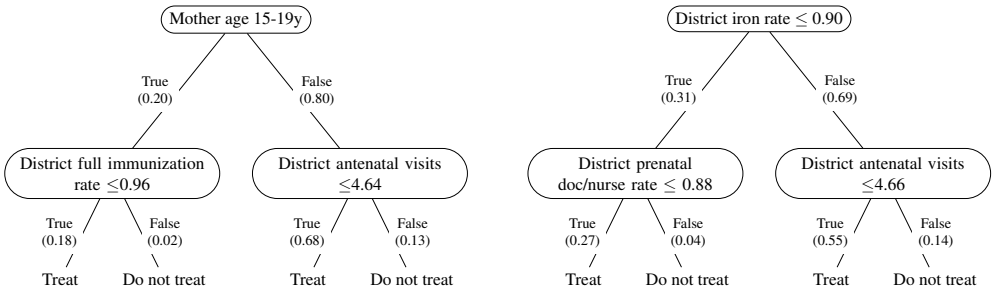
5.2 Lessons for policy targeting

Both our two-way fixed effects model and the causal forest analysis confirm that children born at home are likely to benefit more from CHX cord care, in line with the WHO policy recommendations. Inspired by the approach developed by Kitagawa and Tetenov (2018), we now ask what the optimal targeting policy is

according to the data. Concretely we use the estimates of individual treatment effects from the causal forest to find the optimal policy. A policy consists of a decision rule based on covariates and the optimal policy is defined as the policy that maximizes treatment benefits compared to a random allocation of treatment. In Figure 7 we show two optimal policies that we derive based on the policy learning algorithm developed by Athey and Wager (2021) (and implemented in R with the policytree function due to Sverdrup et al., 2020).

A first policy, where we allow the algorithm to select optimally who should be treated based on the full set of individual and district level variables, is shown in Figure 7a. The resulting optimal targeting depends on maternal age, district level immunization rates and district level average number of antenatal visits. For the second policy, we only allow targeting based on district-level variables. As shown in Figure 7b, the resulting optimal assignment policy is based on the district share of women receiving iron tablets use during pregnancy, the district level share of pregnancies receiving antenatal care from a doctor or nurse, and the district level average number of antenatal visits. Notably, however, the optimal policy does not simply target by baseline neonatal mortality rates, which indicates that the heterogeneity patterns captured in the ML analysis relate to susceptibility to cord infection specifically. While district-level indicators of care quality are correlated, they are not so correlated as to being interchangeable. In particular, the largest coefficient of correlation in absolute terms between a variable selected to define our optimal policy and another variable capturing a different aspect of health care is 0.68 (for iron tablets coverage, which has a 0.68 coefficient of correlation (ρ) with receiving four or more antenatal visits, as shown in Figure A.3). And the selected variable which is most correlated with home delivery is the district average number of antenatal care visits ($\rho = -0.67$).

In Table 7 we compare the two optimal policies in Figure 7 to the WHO policy (of treating only — here, predicted — home births in districts with NMR above 3 percentage points). The data-driven optimal policies both treat between 82 and 86 percent of the sample compared to only 32 percent of the sample for the WHO policy. As Figure 6 shows that there are significant benefits of treatment for at least 50 percent of the population, it is not surprising that the optimal policies with no further constraint than maximizing NMR reductions treat a larger share of the population and are able to reduce neonatal mortality



(a) Individual & district variables **(b)** District variables only

Figure 7: Optimal Policies

Notes: Figure 7a shows the optimal policy based on all variables used in the causal forest (except fixed effects) and all district level variables: antenatal visits (timing and number), iron treatments, tetanus protection, place of delivery, postnatal visits, immunization rate, neonatal mortality rate, nurse/doc delivery support, small baby. Figure 7b shows the optimal policy based on all district level variables. Population shares are shown in parentheses.

rate by more than the WHO policy. The WHO policy would reduce the neonatal mortality by 1.2 percentage point overall, whereas the optimal policies would reduce neonatal mortality by 2.4 percentage points. Interestingly, the benefit of using all variables to design the optimal policy is rather limited, compared to only using district-level variables which are more readily available to policy-makers.

Our — so far — unconstrained policies, which treat a much larger share of the population, would however also be more expensive than the WHO recommended one. We therefore also show results from a constrained optimization in Table 7. The resulting constrained policies are based on an algorithm where we slowly increase the "price" of the treatment (set to zero in the default, unconstrained case) until the number of treated is below or equal the number of treated with the WHO policy. As the results show, the benefits are considerably lower than with the unconstrained policies and only slightly larger than with the WHO policy. Focusing on the more feasible policy targeting only based on district characteristics, the reduction in neonatal mortality would be 1.4 percentage points, compared to the 1.2 percentage point reduction achieved with the WHO policy.

Importantly, the optimal policy is not just to treat everyone even if the treatment had a zero direct monetary cost. This would lead to a reduction in neonatal

Table 7: Reduced mortality by the WHO policy and by optimal policies

	ATT	ATU	%treated	ΔNMR
A. Pre-defined policies				
WHO policy	-0.038***	-0.013***	32.3	-0.012
	(0.007)	(0.004)		
B. Unconstrained optimal policies				
Individual & district variables	-0.028***	0.020	85.7	-0.024
	(0.004)	(0.018)		
District variables only	-0.029***	0.013	82.3	-0.024
	(0.004)	(0.015)		
C. Constrained optimal policies				
Individual & district variables	-0.051***	-0.008*	31.7	-0.016
	(0.009)	(0.004)		
District variables only	-0.048***	-0.012***	28.4	-0.014
	(0.008)	(0.004)		

Notes: Individual & district variables shows the reduction in NMR using the optimal policy based on all variables used in the causal forest (except fixed effects) and district level variables: antenatal visits (timing and number), iron treatments, tetanus protection, place of delivery, post-natal visits, immunization rate, neonatal mortality rate, nurse/doc delivery support, small baby. District variables only shows the reduction in NMR using the optimal policy based on shows the optimal policy based on all district level variables. The column "unconstrained" shows the unconstrained optimal policy. In the column "constrained" we show the results from adding a cost to the treatment until the number of treated individuals is below the WHO policy. Standard errors in parenthesis. The share of the population treated is shown in square brackets. Asterisks indicate significance at the following levels * $p<0.1$, ** $p<0.05$, and *** $p<0.01$.

mortality of 2 percentage points (Table 5, Panel B) instead of the 2.4 percentage points showed in Table 7. This highlights the rare- but important risk that CHX cord care, by departing from the simple public health message of keeping the cord area dry and clean, might have negative externalities such as diverting medical personnel attention from other tasks and sending confused messages about appropriate cord care. Indeed, the CATE is positive for 2.6% of newborns.

Taken together, our optimal policy results show that the WHO guidelines do an excellent job at targeting the third or so of newborns who would stand to benefit the most from CHX, but also exclude many newborns whose chance of survival would be much improved by CHX cord care.

5.3 Reconciling Heterogeneity in RCT Studies

The results of the causal forest suggest that there is significant treatment effect heterogeneity, which might explain that CHX trials were very successful in reducing NMR in three cases, but had no significant effect in two other ones. To assess whether the treatment effect heterogeneity we identify in our national Nepalese sample could explain these different experimental findings across countries, we constructed DHS samples for each of the five subnational regions and time periods in which the RCTs were implemented. We then trained a simplified causal forest in our nationally representative Nepalese dataset based on variables that we are able to observe in all five samples to predict the CATEs for each RCT setting.[18]

Table 8 shows the average CATEs accross the five samples. We observe substantial average treatment effects in the three samples corresponding to areas where RCTs found that CHX significantly reduced neonatal mortality. Namely in Bangladesh (-1.7 percentage points), Nepal (-2.1 percentage points), and Pakistan (-2.5 percentage points). Moreover, for the two samples corresponding to the regions where the RCTs show no significant effects of CHX cord care programs, the average treatment effects are markedly smaller: Tanzania (-0.6 percentage points) and Zambia (-0.5 percentage points). This pattern is further depicted in the histograms of CATEs shown in Figure 8, which also show that there is substantial predicted heterogeneity within each subsample too, with sizeable shares of predicted zero- and even positive treatment effects observed only in the Tanzanian and Zambian subsamples.

Table 6 shows that the first- and fourth quartiles in terms of CATEs differed significantly in terms of place of delivery and maternal education. Table 8 therefore also reports sample averages for these variables across the five samples, and as expected the five settings differ greatly in terms of these variables. Predicted home delivery however only accounts for part of the predicted heterogeneity. The magnitude of the predicted CATE is, for instance, larger in the Nepalese subsample than in the Bangladeshi one despite having a lower rate of home de-

[18] We estimate a causal forest using the full roll-out in Nepal and the same orthogonalization as in the main results described above. However, to make the causal forest comparable across the five samples, the forest is estimated on a reduced set of variables consisting of birth order, gender, maternal age, rural, maternal education, predicted place of delivery, and wealth quintile, as well as 14 district level variables observed in all samples. The fit and results for this forest is shown in column (2) of Appendix Table B.2.

Table 8: Reconciling heterogeneity in RCT studies: average predicted CATEs and key variables across samples

	Bangladesh 2007-2009	Nepal 2002-2005	Pakistan 2007	Tanzania 2011-2014	Zambia 2011-2013
CATE	-0.017	-0.021	-0.025	-0.005	-0.006
Neonatal mortality	0.033	0.051	0.044	0.018	0.021
Predicted home delivery	0.942	0.642	0.909	0.240	0.061
Education: none	0.327	0.854	0.873	0.406	0.060
Education: primary	0.352	0.073	0.079	0.256	0.538
Education: secondary	0.298	0.058	0.048	0.338	0.354
Education: higher	0.023	0.015	0.000	0.000	0.048
Observations	640	137	252	488	867

Notes: This table shows the average CATE based on predictions using the causal forest estimated on full sample of the national roll-out of CHX in Nepal using a reduced set of variables as shown in Appendix X. The table also show average neonatal mortality, predicted home deliveries and education levels. The samples are based on the DHS from Bangladesh (rural areas of Sylhet), Nepal (Sarlahi district), Pakistan (rural areas of Sindh), Tanzania (Pemba Island), and Zambia (Southern Province).

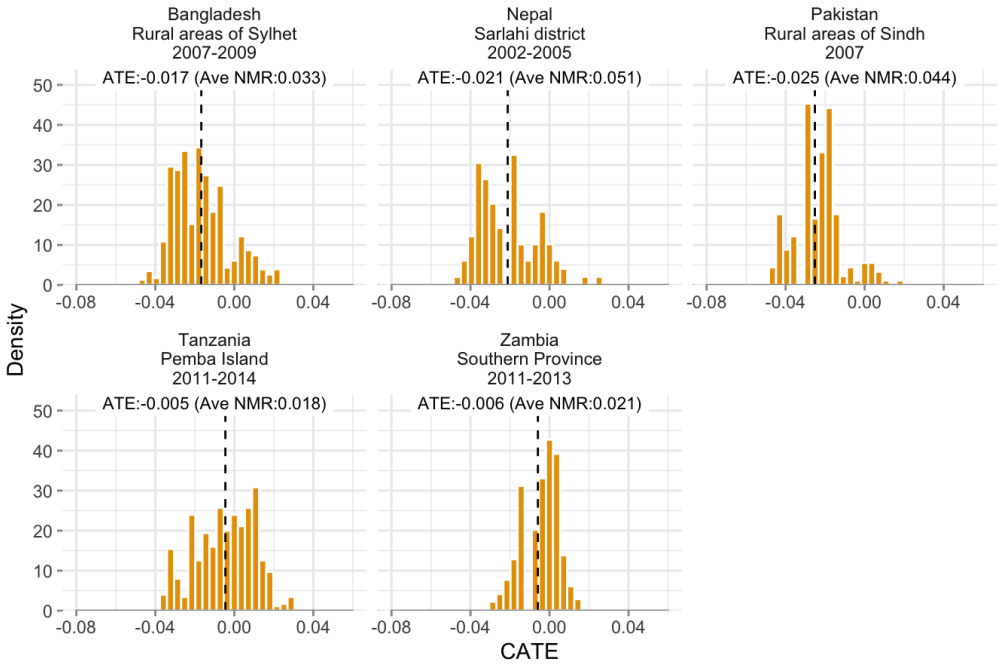


Figure 8: Distribution of predicted CATEs across DHS samples matching the RCT sites

liveries, and the magnitude of the predicted CATE in the Tanzanian subsample is no larger than that of the Zambian subsample despite having a four-fold larger share of predicted home deliveries. Similarly, there is a correlation between overall neonatal mortality and the predicted CATE, but there is no one-to-one correspondence between the two (e.g., the CATE is larger in Pakistan than Nepal despite a lower neonatal mortality rate).

6 Conclusion

Neonatal mortality is an increasingly large contributor to early life mortality across the world, accounting for 45% of under-5 deaths in 2015 compared to 35% in 1980 (Wang et al., 2016). While more efforts and resources than ever before are being targeted at reducing neonatal mortality (Shiffman, 2010) and most neonatal deaths are believed to be preventable at comparatively low cost (Bhutta et al., 2014), there is a wide gap between recommended- and actual practices in low-income countries (Friberg et al., 2010; Requejo et al., 2015). One example of such gap is CHX cord care, for which heterogeneous findings across randomized trials have also led experts to question its effectiveness at scale.

In this paper, we estimate the effect of implementing a nationwide program training health personnel including community health workers to apply CHX to the umbilical stump and to distribute a single CHX dose to mothers who plan to deliver their baby at home. We find that the program led to a large reduction in neonatal mortality (43 percent), driven by reduced neonatal mortality among babies predicted to have been born at home. This provides novel evidence of the effectiveness of CHX cord care outside an experimental setting, and one of the rare instances of any successful nationwide intervention targeting neonatal mortality in a low-income country.

Using recently developed ML techniques, we find evidence of substantial heterogeneity in treatment effects in our nationally representative Nepalese observational data. While place of delivery and average neonatal mortality are good proxies for large treatment effects, the optimal targeting we identify implies treating more than two-and-a-half times more births than the current WHO recommendation based on these two variables. The optimal policy in terms of reducing neonatal mortality is however not to simply treat every birth due to

predicted adverse treatment effects.

Finally, using the conditional treatment effects obtained in our national Nepalese sample, we successfully predict the heterogeneous pattern of results found in the existing experimental literature from five different countries.

References

AL-UBAYDLI, O., J. A. LIST, AND D. L. SUSKIND (2017): "What can we learn from experiments? Understanding the threats to the scalability of experimental results," *American Economic Review*, 107, 282–86.

ARULAMPALAM, W., V. CORRADI, AND D. GUTKNECHT (2017): "Modeling Heaped Duration Data: An Application to Neonatal Mortality," *Journal of Econometrics*, 200, 363–377.

ATHEY, S. AND G. IMBENS (2016): "Recursive partitioning for heterogeneous causal effects," *Proceedings of the National Academy of Sciences*, 113, 7353–7360.

ATHEY, S., J. TIBSHIRANI, S. WAGER, ET AL. (2019): "Generalized random forests," *Annals of Statistics*, 47, 1148–1178.

ATHEY, S. AND S. WAGER (2021): "Policy learning with observational data," *Econometrica*, 89, 133–161.

BARHAM, T. (2011): "A healthier start: The effect of conditional cash transfers on neonatal and infant mortality in rural Mexico," *Journal of Development Economics*, 94, 74–85.

BAUERNSCHUSTER, S., A. DRIVA, AND E. HORNUNG (2017): "Bismarck's Health Insurance and the Mortality Decline," *Journal of the European Economic Association*.

BECKETT, M., J. DA VANZO, N. SASTRY, C. PANIS, AND C. PETERSON (2001): "The Quality of Retrospective Data: An Examination of Long-Term Recall in a Developing Country," *Journal of Human Resources*, 593–625.

BENNETT, J., J. MACIA, H. TRAVERSO, S. BANOAGHA, C. MALOOLY, AND J. BORING (1997): "Protective effects of topical antimicrobials against neonatal tetanus." *International journal of epidemiology*, 26, 897–903.

BHALOTRA, S. R., R. ROCHA, AND R. R. SOARES (2019): "Does Universalization of Healthwork? Evidence from Health Systems Restructuring and Expansion in Brazil," *IZA Discussion Paper #12111*.

BHUTTA, Z. A., J. K. DAS, R. BAHL, J. E. LAWN, R. A. SALAM, V. K. PAUL, M. J. SANKAR, H. BLENCOWE, A. RIZVI, V. B. CHOU, ET AL. (2014): "Can available interventions end preventable deaths in mothers, newborn babies, and stillbirths, and at what cost?" *The Lancet*, 384, 347–370.

BRUHN, J. (2020): "The Consequences of Sorting for Understanding School Quality," Tech. rep.

CHERNOZHUKOV, V., M. DEMIRER, E. DUFLO, AND I. FERNÁNDEZ-VAL (2020): "Generic Machine Learning Inference on Heterogenous Treatment Effects in Randomized Experiments," .

CONTI, G. AND R. GINJA (forthcoming): "Health Insurance and Child Health: Evidence from Mexico," *Journal of Human Resources*.

DE CHAISEMARTIN, C. AND X. D'HAULTFOEUILLE (2020): "Two-way Fixed Effects Estimators with Heterogeneous Treatment Effects," *American Economic Review*, 110, 2964–96.

DEPARTMENT OF HEALTH SERVICES (2015): *Annual Report 2014/15*, Kathmandu.

EL ARIFEEN, S., L. C. MULLANY, R. SHAH, I. MANNAN, S. M. RAHMAN, M. R. R. TALUKDER, N. BEGUM, A. AL-KABIR, G. L. DARMSTADT, M. SANTOSHAM, ET AL. (2012): "The effect of cord cleansing with chlorhexidine on neonatal mortality in rural Bangladesh: a community-based, cluster-randomised trial," *The Lancet*, 379, 1022–1028.

FITZPATRICK, A. (2018): "The Price of Labor: Evaluating the Impact of Eliminating User Fees on Maternal and Infant Health Outcomes," in *AEA Papers and Proceedings*, vol. 108, 412–15.

FRIBERG, I. K., M. V. KINNEY, J. E. LAWN, K. J. KERBER, M. O. ODUBANJO, A.-M. BERGH, N. WALKER, E. WEISSMAN, M. CHOPRA, R. E. BLACK, ET AL. (2010): "Sub-Saharan Africa's Mothers, Newborns, and Children: How Many Lives Could Be Saved with Targeted Health Interventions?" *PLoS Med*, 7, e1000295.

GOODMAN-BACON, A. (2021): "Difference-in-Differences with Variation in Treatment Timing," *Journal of Econometrics*.

HIMS (2014): *HIMS Database*, Health Information Management System, Nepal.

HODGINS, S., L. KHANAL, N. JOSHI, S. PENFOLD, S. TULADHAR, P. R. SHRESTHA, B. LAMICHHANE, P. DAWSON, T. GUENTHER, S. SINGH, ET AL. (2019): "Achieving and sustaining impact at scale for a newborn intervention in Nepal: a mixed-methods study," *Journal of Global Health Reports*, 3.

HODGINS, S., Y. PRADHAN, L. KHANAL, S. UPRETI, AND N. P. KC (2013): "Chlorhexidine for Umbilical Cord Care: Game-Changer for Newborn Survival?" *Global Health: Science and Practice*, 1, 5–10.

IMDAD, A., L. C. MULLANY, A. H. BAQUI, S. EL ARIFEEN, J. M. TIELSCH, S. K. KHATRY, R. SHAH, S. COUSENS, R. E. BLACK, AND Z. A. BHUTTA (2013): "The effect of umbilical cord cleansing with chlorhexidine on omphalitis and neonatal mortality in community settings in developing countries: a meta-analysis," *BMC public health*, 13, S15.

JSI (2017): *Monitoring and Evaluation of the Chlorhexidine "Navi" Care Program Technical Brief #1*.

JSI RESEARCH & TRAINING INSTITUTE (2017): *Use of Chlorhexidine for Cord Care, Social Change Communication, Experience from Nepal*, Kathmandu.

KHANAL, L. (2015): "Institutionalizing Chlorhexidine Program and Maintaining Coverage Chlorhexidine Cord Care Program in Nepal," `https://www.healthynewbornnetwork.org/hnn-content/uploads/Institutionalizing-Chlorhexidine-Program-and-Maintaining-Coverage-in-Nepal.pdf`.

KITAGAWA, T. AND A. TETENOV (2018): "Who should be treated? empirical welfare maximization methods for treatment choice," *Econometrica*, 86, 591–616.

KOWALSKI, A. (forthcoming): "Reconciling Seemingly Contradictory Results from the Oregon Health Insurance Experiment and the Massachusetts Health Reform," *Review of Economics and Statistics*.

KOWALSKI, A. E. (2021): "Mammograms and Mortality: How Has the Evidence Evolved?" *Journal of Economic Perspectives*, 35, 119–40.

LIM, S. S., L. DANDONA, J. A. HOISINGTON, S. L. JAMES, M. C. HOGAN, AND E. GAKIDOU (2010): "India's Janani Suraksha Yojana, a conditional cash transfer programme to increase births in health facilities: an impact evaluation," *The Lancet*, 375, 2009 – 2023.

LIU, L., S. OZA, D. HOGAN, Y. CHU, J. PERIN, J. ZHU, J. E. LAWN, S. COUSENS, C. MATHERS, AND R. E. BLACK (2016): "Global, regional, and national causes of under-5 mortality in 2000–15: an updated systematic analysis with implications for the Sustainable Development Goals," *The Lancet*, 388, 3027–3035.

LÓPEZ-MEDINA, M. D., M. LINARES-ABAD, A. B. LÓPEZ-ARAQUE, AND I. M. LÓPEZ-MEDINA (2019): "Dry care versus chlorhexidine cord care for prevention of omphalitis. Systematic review with meta-analysis," *Revista latino-americana de enfermagem*, 27.

MCKINNON, B., S. HARPER, J. S. KAUFMAN, AND Y. BERGEVIN (2015): "Removing user fees for facility-based delivery services: a difference-in-differences evaluation from ten sub-Saharan African countries," *Health policy and planning*, 30, 432–441.

MEAGER, R. (2019): "Understanding the Average Impact of Microcredit Expansions: A Bayesian Hierarchical Analysis of Seven Randomized Experiments," *American Economic Journal: Applied Economics*, 11, 57–91.

MINISTRY OF HEALTH [NEPAL] AND NEW ERA AND ICF (2017a): *Nepal Demographic and Health Survey 2016*, Ministry of Health [Nepal].

——— (2017b): *Nepal Demographic and Health Survey 2016 [Dataset]. NPBR7HFL.DTA*, Ministry of Health [Nepal], `https://dhsprogram.com/data/dataset/Nepal_Standard-DHS_2016.cfm?flag=0`. Last accessed July 11 2018.

MULLANY, L. C., G. L. DARMSTADT, S. K. KHATRY, J. KATZ, S. C. LECLERQ, S. SHRESTHA, R. ADHIKARI, AND J. M. TIELSCH (2006): "Topical applications of chlorhexidine to the umbilical cord for prevention

of omphalitis and neonatal mortality in southern Nepal: a community-based, cluster-randomised trial," *The Lancet*, 367, 910–918.

NEWHOUSE, J. P. (2021): "An Ounce of Prevention," *Journal of Economic Perspectives*, 35, 101–18.

OSRIN, D. AND T. COLBOURN (2016): "No reason to change WHO guidelines on cleansing the umbilical cord," *The Lancet Global Health*, 4, e766–e768.

PAINTER, K. AND J. FELDMAN (2019): *Omphalitis*, `https://www.ncbi.nlm.nih.gov/books/NBK513338/`.

PATH (2017): *From Research to Use: Saving Newborn Lives With Chlorhexidine for Umbilical Cord Care*.

PHILIBERT, A., M. RAVIT, V. RIDDE, I. DOSSA, E. BONNET, F. BEDECARRATS, AND A. DUMONT (2017): "Maternal and neonatal health impact of obstetrical risk insurance scheme in Mauritania: a quasi experimental before-and-after study," *Health policy and planning*, 32, 405–417.

PONCE HARDY, V. (2018): "Chlorhexidine for neonatal infection: the 'game-changer' we've all been waiting for?" *Tropical Medicine & International Health*, 23, 252–253.

POWELL-JACKSON, T., S. MAZUMDAR, AND A. MILLS (2015): "Financial incentives in health: New evidence from India's Janani Suraksha Yojana," *Journal of health economics*, 43, 154–169.

REQUEJO, J. H., J. BRYCE, A. J. BARROS, P. BERMAN, Z. BHUTTA, M. CHOPRA, B. DAELMANS, A. DE FRANCISCO, J. LAWN, B. MALIQI, ET AL. (2015): "Countdown to 2015 and Beyond: Fulfilling the Health Agenda for Women and Children," *The Lancet*, 385, 466–476.

SANKAR, M., A. CHANDRASEKARAN, A. RAVINDRANATH, R. AGARWAL, AND V. PAUL (2016): "Umbilical cord cleansing with chlorhexidine in neonates: a systematic review," *Journal of Perinatology*, 36, S12–S20.

SAZAWAL, S., U. DHINGRA, S. M. ALI, A. DUTTA, S. DEB, S. M. AME, M. H. MKASHA, A. YADAV, AND R. E. BLACK (2016): "Efficacy of

chlorhexidine application to umbilical cord on neonatal mortality in Pemba, Tanzania: a community-based randomised controlled trial," *The Lancet Global Health*, 4, e837–e844.

SEMRAU, K. E., J. HERLIHY, C. GROGAN, K. MUSOKOTWANE, K. YEBOAH-ANTWI, R. MBEWE, B. BANDA, C. MPAMBA, F. HAMOMBA, P. PILINGANA, ET AL. (2016): "Effectiveness of 4% chlorhexidine umbilical cord care on neonatal mortality in Southern Province, Zambia (ZamCAT): a cluster-randomised controlled trial," *The Lancet Global Health*, 4, e827–e836.

SHIFFMAN, J. (2010): "Issue Attention in Global Health: The Case of Newborn Survival," *The Lancet*, 375, 2045–2049.

SOOFI, S., S. COUSENS, A. IMDAD, N. BHUTTO, N. ALI, AND Z. A. BHUTTA (2012): "Topical application of chlorhexidine to neonatal umbilical cords for prevention of omphalitis and neonatal mortality in a rural district of Pakistan: a community-based, cluster-randomised trial," *The Lancet*, 379, 1029–1036.

SVERDRUP, E., A. KANODIA, Z. ZHOU, S. ATHEY, AND S. WAGER (2020): "policytree: Policy learning via doubly robust empirical welfare maximization over trees," *Journal of Open Source Software*, 5, 2232.

TIBSHIRANI, J., S. ATHEY, R. FRIEDBERG, V. HADAD, D. HIRSHBERG, L. MINER, E. SVERDRUP, S. WAGER, AND M. WRIGHT (2021): *grf: Generalized Random Forests*, r package version 2.0.2.

VAN DE POEL, E., G. FLORES, P. IR, AND O. O'DONNELL (2016): "Impact of performance-based financing in a low-resource setting: a decade of experience in Cambodia," *Health Economics*, 25, 688–705.

VARIAN, H. R. (2014): "Big data: New tricks for econometrics," *Journal of Economic Perspectives*, 28, 3–28.

VIVALT, E. (2020): "How Much Can We Generalize from Impact Evaluations?" *Journal of the European Economics Association*, 18, 3045–3089.

WANG, H., Z. A. BHUTTA, M. M. COATES, M. COGGESHALL, L. DANDONA, K. DIALLO, E. B. FRANCA, M. FRASER, N. FULLMAN, P. W. GETHING, ET AL. (2016): "Global, regional, national, and selected subnational levels of stillbirths, neonatal, infant, and under-5 mortality, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015," *The Lancet*, 388, 1725–1774.

WHO (2020): "Global Health Observatory," https://www.who.int/data/gho/data/indicators/indicator-details/GHO/neonatal-mortality-rate-(deaths-per-1000-live-births), accessed: 2020-12-08.

WORLD HEALTH ORGANIZATION (2015): "Postnatal care for mothers and newborns: Highlights from the World Health Organization 2013 Guidelines," *Available from: http://www.who.int/maternal_child_adolescent/publications/WHOMCA-PNC-2014-Briefer_A*, 4.

——— (2020): "Advice on the Use of Masks in the Context of COVID-19: Interim Guidance (5 June 2020)," Tech. rep.

A Appendix (for Online Publication Only)

Table A.1: Effect of CHX-NCP on Home delivery and Antenatal Care

	(1)	(2)	(3)	(4)
A. Dependent variable: home delivery (binary; mean: 0.41)				
Program: CHX	-0.020	-0.014	-0.028	-0.032
	(0.027)	(0.024)	(0.024)	(0.025)
Observations	4955	4955	4955	4955
B. Dependent variable: antenatal visits (count; mean: 4.23)				
Program: CHX	0.001	-0.028	0.052	0.071
	(0.141)	(0.133)	(0.117)	(0.109)
Observations	3966	3966	3966	3966
C. Dependent variable: antenatal visits above median (binary; mean: 0.37)				
Program: CHX	-0.012	-0.016	-0.004	0.001
	(0.030)	(0.029)	(0.029)	(0.030)
Observations	3966	3966	3966	3966
DEM controls	No	Yes	Yes	Yes
SES controls	No	No	Yes	Yes
Program controls	No	No	No	Yes

Notes: All specifications are estimated as linear probability models using OLS with district and month of birth fixed effects. Demographic controls include birth order (three indicators), five year maternal age group indicators, and gender. SES controls include education (three indicators), wealth (four indicators), rural indicator, altitude quintile indicators, and a Dalit ethnicity indicator. Program controls include controls for the CB-NCP and CB-IMNCI health programs. The place of delivery is predicted using the linear probability shown in Appendix Table A.2. Bootstrapped standard errors clustered at the district level in parentheses. Asterisks indicate significance at the following levels * p<0.1, ** p<0.05, and *** p<0.01.

Table A.2: Predicting home deliveries

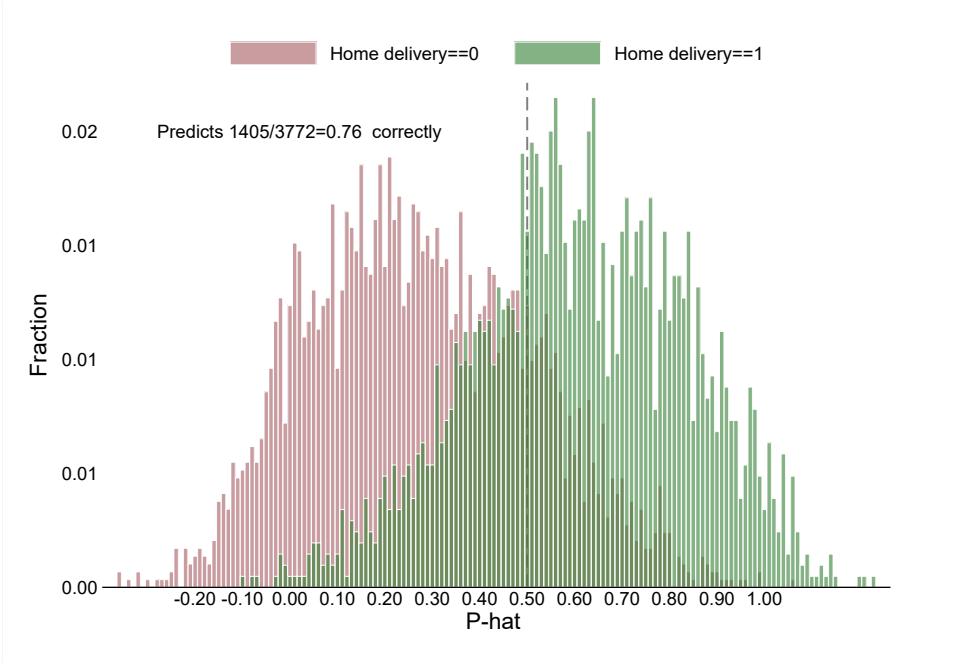
	Logit (1)	LPM (2)
Female	0.004	0.006
	(0.012)	(0.012)
First born	-0.248***	-0.267***
	(0.026)	(0.027)
Second born	-0.090***	-0.112***
	(0.019)	(0.022)
Third born	-0.022	-0.029
	(0.017)	(0.019)
Mother age 15-19y	0.069	0.067
	(0.062)	(0.065)
Mother age 20-24y	0.034	0.039
	(0.059)	(0.063)
Mother age 25-29y	0.015	0.018
	(0.057)	(0.062)
Mother age 30-34y	-0.048	-0.046
	(0.059)	(0.064)
Mother age 35-39y	-0.040	-0.048
	(0.060)	(0.064)
Ethnicity: hill chhetri	0.042	0.031
	(0.039)	(0.032)
Ethnicity: terai brahmin/chhetri	-0.091	-0.126**
	(0.066)	(0.055)
Ethnicity: other terai caste	0.065	0.040
	(0.047)	(0.045)
Ethnicity: hill dalit	0.053	0.034
	(0.046)	(0.043)
Ethnicity: terai dalit	0.095*	0.092*
	(0.050)	(0.048)
Ethnicity: newar	0.057	0.044
	(0.056)	(0.043)
Ethnicity: hill janajati	0.112***	0.099***
	(0.039)	(0.033)
Ethnicity: terai janajati	0.031	-0.000
	(0.047)	(0.038)
Ethnicity: muslim	0.063	0.027
	(0.050)	(0.048)
Ethnicity: other	-0.000	0.071
	(0.137)	(0.066)
Rural	0.120***	0.130***
	(0.025)	(0.029)
Altitude in 1st quintile	-0.139*	-0.108
	(0.071)	(0.077)
Altitude in 2nd quintile	-0.167**	-0.141*
	(0.074)	(0.081)
Altitude in 3rd quintile	-0.114**	-0.093

Continued on next page

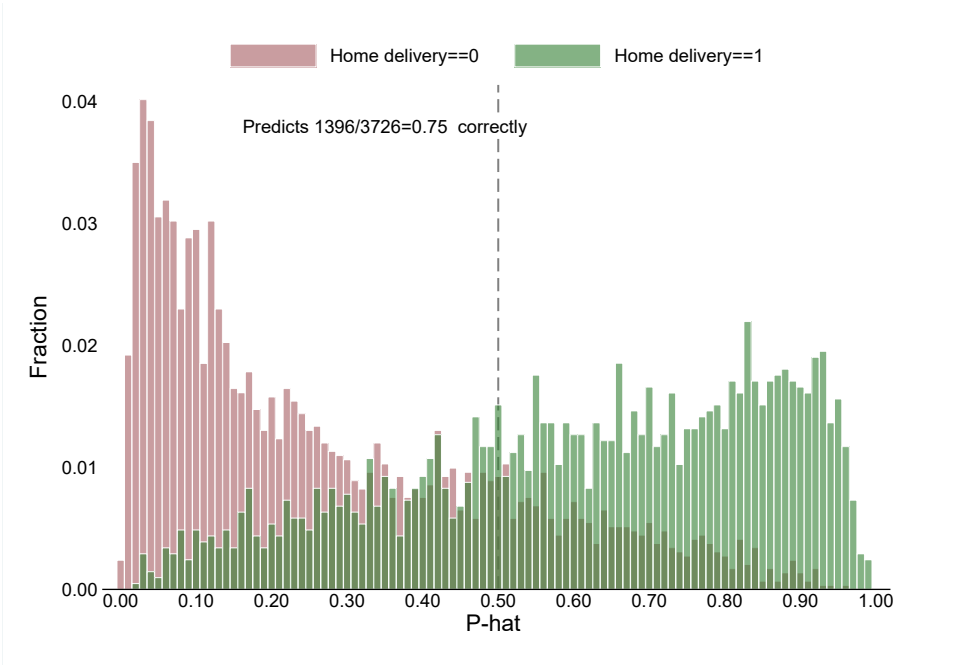
Continued from previous page

	Logit (1)	LPM (2)
	(0.056)	(0.064)
Altitude in 4th quintile	0.014	0.012
	(0.048)	(0.054)
Education: no education	0.146***	0.137***
	(0.030)	(0.027)
Education: primary	0.104***	0.082***
	(0.030)	(0.026)
Education: secondary	0.071***	0.035*
	(0.027)	(0.020)
Wealth in 1st quintile	0.288***	0.278***
	(0.034)	(0.033)
Wealth in 2nd quintile	0.235***	0.208***
	(0.032)	(0.029)
Wealth in 3rd quintile	0.119***	0.084***
	(0.030)	(0.027)
Wealth in 4th quintile	0.090***	0.051*
	(0.032)	(0.026)
Observations	4,956	4,956
P(home birth)>0.5|Home birth==1	1396	1405
P(home birth)<0.5|Home birth==0	2330	2367
P(home birth)>0.5|Home birth==0	545	543
P(home birth)<0.5|Home birth==1	640	641
Correct predictions (share)	0.752	0.761

Notes: Column (1) shows average marginal effects from estimating a Logit specification. Column (2) shows point estimates from estimating a linear probability models. Both regressions include district fixed effects and date of birth, defined by Nepali month and year of birth, fixed effects. Bootstrapped standard errors based on 200 iterations and clustered at the district level in parentheses. Asterisks indicate significance at the following levels * p<0.1, ** p<0.05, and *** p<0.01.



(a) LPM



(b) LOGIT

Figure A.1: Distribution of predicted probabilities for home births, by actual home birth (see Table A.2 for details).

Table A.3: Balancing table. Dependent variable: CHX.

	All	P(home birth) <0.5	P(home birth) >0.5
	(1)	(2)	(3)
Female	0.000	0.003	-0.002
	(0.002)	(0.003)	(0.003)
Second born	-0.004	-0.006	0.004
	(0.004)	(0.006)	(0.006)
Third born	-0.007	-0.021***	0.005
	(0.005)	(0.008)	(0.006)
Parity four or higher	-0.001	-0.014	0.005
	(0.005)	(0.010)	(0.007)
Mother age 20-24y	-0.002	-0.004	0.001
	(0.004)	(0.005)	(0.005)
Mother age 25-29y	0.000	0.004	0.004
	(0.004)	(0.007)	(0.006)
Mother age 30-34y	-0.002	0.002	0.005
	(0.007)	(0.011)	(0.007)
Mother age 35-39y	-0.000	-0.010	0.009
	(0.011)	(0.024)	(0.012)
Mother age 40-45y	0.026	0.101	-0.005
	(0.027)	(0.078)	(0.028)
Ethnicity: hill chhetri	-0.005	-0.008	0.000
	(0.006)	(0.008)	(0.009)
Ethnicity: terai brahmin/chhetri	-0.003	-0.016	-0.019
	(0.010)	(0.021)	(0.022)
Ethnicity: other terai caste	-0.006	-0.001	-0.009
	(0.008)	(0.013)	(0.013)
Ethnicity: hill dalit	0.006	0.009	0.006
	(0.006)	(0.010)	(0.008)
Ethnicity: terai dalit	0.006	-0.001	-0.001
	(0.009)	(0.029)	(0.012)
Ethnicity: newar	0.007	0.015	-0.004
	(0.010)	(0.017)	(0.015)
Ethnicity: hill janajati	0.000	0.004	0.005
	(0.006)	(0.011)	(0.008)
Ethnicity: terai janajati	0.001	0.010	-0.011
	(0.007)	(0.010)	(0.014)
Ethnicity: muslim	-0.008	-0.031	-0.003
	(0.007)	(0.021)	(0.014)
Ethnicity: other	0.052**	0.066**	-0.052
	(0.025)	(0.033)	(0.071)
Rural	-0.007*	-0.020**	-0.000
	(0.004)	(0.008)	(0.004)
Altitude in 1st quintile	0.013	0.012	0.016
	(0.011)	(0.026)	(0.013)
Altitude in 2nd quintile	0.003	0.002	0.011
	(0.010)	(0.023)	(0.013)

Continued on next page

Continued from previous page

	All	P(home birth) <0.5	P(home birth) >0.5
	(1)	(2)	(3)
Altitude in 3rd quintile	-0.007	-0.013	0.009
	(0.007)	(0.016)	(0.010)
Altitude in 4th quintile	-0.000	0.006	0.009
	(0.007)	(0.015)	(0.007)
Education: primary	0.005	0.009	-0.003
	(0.005)	(0.007)	(0.005)
Education: secondary	-0.002	0.008	-0.016*
	(0.007)	(0.009)	(0.010)
Education: higher	-0.016	0.000	-0.022
	(0.011)	(0.011)	(0.021)
Wealth in 2nd quintile	0.007	-0.004	0.009*
	(0.004)	(0.010)	(0.005)
Wealth in 3rd quintile	0.002	-0.008	0.000
	(0.005)	(0.013)	(0.007)
Wealth in 4th quintile	0.001	-0.010	0.004
	(0.005)	(0.013)	(0.007)
Wealth in 5th quintile	0.005	-0.008	-0.014
	(0.006)	(0.012)	(0.011)
Program: CB-NCP	0.373***	0.374***	0.374***
	(0.069)	(0.085)	(0.072)
Program: CB-IMNCI	-0.063	-0.065	-0.075
	(0.079)	(0.094)	(0.088)
Constant	0.082***	0.108***	0.049**
	(0.019)	(0.033)	(0.019)
Observations	23465	23465	23465

Notes: All specifications are estimated with district and month of birth fixed effects. The place of delivery is predicted using the linear probability shown in Appendix Table A.2. Bootstrapped standard errors clustered at the district level in parentheses. Asterisks indicate significance at the following levels * p<0.1, ** p<0.05, and *** p<0.01.

Table A.4: Regression results: The effect of CHX-NCP on neonatal mortality - Dependent variable: Mortality by ≤1m. Reporting all coefficient estimates

	Sample			
	All	All	P(home birth) <0.5	P(home birth) >0.5
	(1)	(2)	(3)	(4)
Female	-0.014***	-0.014***	-0.007*	-0.021***
	(0.003)	(0.003)	(0.004)	(0.005)
First born	0.003	0.001	0.011	-0.003
	(0.005)	(0.005)	(0.009)	(0.009)
Second born	-0.007	-0.008*	0.004	-0.016**
	(0.005)	(0.005)	(0.010)	(0.007)
Third born	-0.008*	-0.008*	-0.002	-0.010
	(0.005)	(0.005)	(0.008)	(0.006)
Mother age 15-19y	0.019	0.019	0.034**	0.022
	(0.022)	(0.022)	(0.017)	(0.026)
Mother age 20-24y	-0.003	-0.003	0.015	-0.005
	(0.020)	(0.020)	(0.016)	(0.024)
Mother age 25-29y	-0.011	-0.011	0.011	-0.017
	(0.021)	(0.020)	(0.017)	(0.024)
Mother age 30-34y	-0.008	-0.008	0.019	-0.017
	(0.021)	(0.021)	(0.017)	(0.027)
Mother age 35-39y	-0.006	-0.006	0.017	-0.013
	(0.022)	(0.022)	(0.020)	(0.026)
Ethnicity: hill chhetri	-0.005	-0.005	-0.000	-0.016
	(0.005)	(0.005)	(0.005)	(0.012)
Ethnicity: terai brahmin/chhetri	0.002	0.002	0.009	-0.010
	(0.015)	(0.015)	(0.016)	(0.036)
Ethnicity: other terai caste	0.002	0.002	0.006	-0.012
	(0.009)	(0.009)	(0.012)	(0.018)
Ethnicity: hill dalit	-0.004	-0.004	-0.006	-0.010
	(0.006)	(0.006)	(0.007)	(0.013)
Ethnicity: terai dalit	0.019	0.019	0.002	0.008
	(0.012)	(0.012)	(0.018)	(0.018)
Ethnicity: newar	0.003	0.003	0.001	-0.002
	(0.009)	(0.009)	(0.011)	(0.022)
Ethnicity: hill janajati	-0.006	-0.005	-0.010	-0.009
	(0.006)	(0.006)	(0.007)	(0.013)
Ethnicity: terai janajati	0.007	0.006	0.016*	-0.022
	(0.007)	(0.007)	(0.009)	(0.014)
Ethnicity: muslim	-0.006	-0.006	0.002	-0.020
	(0.012)	(0.012)	(0.012)	(0.024)
Ethnicity: other	0.002	0.002	0.003	-0.005
	(0.023)	(0.023)	(0.051)	(0.076)
Rural	0.003	0.003	0.010	-0.001
	(0.003)	(0.004)	(0.006)	(0.005)
Altitude in 1st quintile	-0.014	-0.014	-0.026	-0.012
	(0.012)	(0.012)	(0.019)	(0.023)

Continued on next page

Continued from previous page

	Sample			
	All	All	P(home birth) <0.5	P(home birth) >0.5
	(1)	(2)	(3)	(4)
Altitude in 2nd quintile	-0.026**	-0.026**	-0.028*	-0.032
	(0.011)	(0.011)	(0.016)	(0.022)
Altitude in 3rd quintile	-0.015*	-0.015*	-0.021	-0.013
	(0.008)	(0.008)	(0.015)	(0.011)
Altitude in 4th quintile	-0.007	-0.007	-0.017	-0.006
	(0.007)	(0.007)	(0.012)	(0.009)
Education: no education	0.018***	0.020***	0.023***	0.042***
	(0.005)	(0.005)	(0.006)	(0.013)
Education: primary	0.008*	0.010**	0.012**	0.033***
	(0.005)	(0.005)	(0.006)	(0.012)
Education: secondary	0.007	0.008*	0.005	0.039**
	(0.004)	(0.004)	(0.005)	(0.016)
Wealth in 1st quintile	0.017**	0.018**	0.011	0.019
	(0.007)	(0.008)	(0.011)	(0.019)
Wealth in 2nd quintile	0.019***	0.020***	0.019**	0.018
	(0.006)	(0.006)	(0.008)	(0.019)
Wealth in 3rd quintile	0.013***	0.013***	0.009	0.013
	(0.005)	(0.005)	(0.006)	(0.016)
Wealth in 4th quintile	0.004	0.004	-0.000	0.006
	(0.005)	(0.005)	(0.006)	(0.016)
Program: CB-NCP	0.006	0.005	0.006	-0.000
	(0.006)	(0.006)	(0.008)	(0.008)
Program: CB-IMNCI	-0.003	-0.003	0.002	-0.010
	(0.006)	(0.006)	(0.008)	(0.010)
CHX	-0.018**	-0.007	0.001	-0.028**
	(0.007)	(0.007)	(0.009)	(0.011)
1[P(home birth)>0.5]		-0.001		
		(0.005)		
CHX × 1[P(home birth)>0.5]		-0.021***		
		(0.008)		
CHX + CHX × 1[P(home birth)>0.5]		-0.028***		
		(0.008)		
Observations	23,465	23,465	10,860	12,605
Clusters	73	73	73	73
Control mean of dep. var	0.042	0.042	0.033	0.050
P-val (dif across sample)				0.031

Notes: All specifications are estimated with district and month of birth fixed effects. The place of delivery is predicted using the linear probability shown in Appendix Table A.2. Bootstrapped standard errors clustered at the district level in parentheses. Asterisks indicate significance at the following levels * p<0.1, ** p<0.05, and *** p<0.01.

Table A.5: Placebo regression with 12m lead indicator and pre period only. - Dependent variable: Mortality by $\leq$1m.

	Sample			
	All	All	P(home birth) <0.5	P(home birth) >0.5
	(1)	(2)	(3)	(4)
CHX	-0.005	-0.007	-0.011	0.003
	(0.008)	(0.010)	(0.011)	(0.013)
1[P(home birth)>0.5]		-0.009		
		(0.006)		
CHX $\times$ 1[P(home birth)>0.5]		0.004		
		(0.013)		
CHX + CHX $\times$ 1[P(home birth)>0.5]		-0.003		
		(0.010)		
Observations	20,321	20,321	7,262	13,047
Clusters	73	73	73	73
Control mean of dep. var	0.043	0.043	0.032	0.049
P-val (dif across sample)				0.418

Notes: All specifications are estimated as linear probability models using OLS with the full set of demographic, SES, and program controls. Demographic controls include birth order (three indicators), five year maternal age group indicators, and gender. SES controls include education (three indicators), wealth (four indicators), rural indicator, altitude quintile indicators, and ethnicity indicators. Program controls include controls for the CB-NCP and CB-IMNCI health programs. All specifications are estimated with district and month of birth fixed effects. The place of delivery is predicted using the linear probability shown in Appendix Table A.2. Bootstrapped standard errors clustered at the district level in parentheses. Asterisks indicate significance at the following levels * p<0.1, ** p<0.05, and *** p<0.01.

Table A.6: Regression results including only the first post treatment year. - Dependent variable: Mortality by $\leq$1m.

	Sample			
	All	All	P(home birth)	
			<0.5	>0.5
	(1)	(2)	(3)	(4)
CHX	-0.015	-0.002	0.004	-0.027**
	(0.009)	(0.012)	(0.013)	(0.011)
1[P(home birth)>0.5]		-0.007		
		(0.005)		
CHX × 1[P(home birth)>0.5]		-0.025*		
		(0.013)		
CHX + CHX × 1[P(home birth)>0.5]		-0.027***		
		(0.009)		
Observations	21,185	21,185	9,050	12,129
Clusters	73	73	73	73
Control mean of dep. var	0.042	0.042	0.034	0.049
P-val (dif across sample)				0.060

Notes: All specifications are estimated as linear probability models using OLS with the full set of demographic, SES, and program controls. Demographic controls include birth order (three indicators), five year maternal age group indicators, and gender. SES controls include education (three indicators), wealth (four indicators), rural indicator, altitude quintile indicators, and ethnicity indicators. Program controls include controls for the CB-NCP and CB-IMNCI health programs. All specifications are estimated with district and month of birth fixed effects. The place of delivery is predicted using the linear probability shown in Appendix Table A.2. Bootstrapped standard errors clustered at the district level in parentheses. Asterisks indicate significance at the following levels * p<0.1, ** p<0.05, and *** p<0.01.

Table A.7: Regression results: The effect of CHX-NCP on neonatal mortality. - Dependent variable: Mortality by $\leq$1m. Using actual place of delivery.

	Sample			
	All	All	Home birth	
			No	Yes
	(1)	(2)	(3)	(4)
Program: CHX	-0.003	0.007	0.007	-0.012
	(0.008)	(0.008)	(0.010)	(0.016)
CHX × Home Delivery		-0.023**		
		(0.011)		
Home Delivery		0.020**		
		(0.009)		
CHX + CHX × Home Delivery		-0.016		
P-val (CHX + CHX × Home Delivery=0)		0.148		
Observations	4,839	4,839	2,829	2,008
Clusters	73	73	71	70
Control mean of dep. var	0.025	0.025	0.013	0.041

Notes: All specifications are estimated as linear probability models using OLS with the full set of demographic, SES, and program controls. Demographic controls include birth order (three indicators), five year maternal age group indicators, and gender. SES controls include education (three indicators), wealth (four indicators), rural indicator, altitude quintile indicators, and ethnicity indicators. Program controls include controls for the CB-NCP and CB-IMNCI health programs. All specifications are estimated with district and month of birth fixed effects. Bootstrapped standard errors based on 200 iterations and clustered at the district level in parentheses. Asterisks indicate significance at the following levels * p<0.1, ** p<0.05, and *** p<0.01.

Table A.8: Regression results - Dependent variable: Mortality by <1m.

	Sample			
	All	All	P(home birth) <0.5	P(home birth) >0.5
	(1)	(2)	(3)	(4)
CHX	-0.014**	-0.006	0.002	-0.022**
	(0.007)	(0.007)	(0.008)	(0.010)
1[P(home birth)>0.5]		-0.002		
		(0.005)		
CHX × 1[P(home birth)>0.5]		-0.016**		
		(0.008)		
CHX + CHX × 1[P(home birth)>0.5]		-0.022***		
		(0.008)		
Observations	23,552	23,552	10,920	12,631
Clusters	73	73	73	73
Control mean of dep. var	0.037	0.037	0.028	0.044
P-val (dif across sample)				0.057

Notes: All specifications are estimated as linear probability models using OLS with the full set of demographic, SES, and program controls. Demographic controls include birth order (three indicators), five year maternal age group indicators, and gender. SES controls include education (three indicators), wealth (four indicators), rural indicator, altitude quintile indicators, and ethnicity indicators. Program controls include controls for the CB-NCP and CB-IMNCI health programs. The place of delivery is predicted using the linear probability shown in Appendix Table A.2. All specifications are estimated with district and month of birth fixed effects. Bootstrapped standard errors clustered at the district level in parentheses. Asterisks indicate significance at the following levels * p<0.1, ** p<0.05, and *** p<0.01.

Table A.9: Regression results with survey weights: The effect of CHX-NCP on neonatal mortality. - Dependent variable: Mortality by ≤1m.

	Sample			
	All	All	P(home birth) <0.5	P(home birth) >0.5
	(1)	(2)	(3)	(4)
CHX	-0.017**	-0.004	0.001	-0.027**
	(0.007)	(0.008)	(0.010)	(0.012)
1[P(home birth)>0.5]		0.007		
		(0.006)		
CHX × 1[P(home birth)>0.5]		-0.025***		
		(0.009)		
CHX + CHX × 1[P(home birth)>0.5]		-0.029***		
		(0.009)		
Observations	23,465	23,465	10,966	12,498
Clusters	73	73	73	73
Control mean of dep. var	0.042	0.042	0.033	0.051
P-val (dif across sample)				0.084

Notes: All specifications are estimated as linear probability models using OLS with the full set of demographic, SES, and program controls. Demographic controls include birth order (three indicators), five year maternal age group indicators, and gender. SES controls include education (three indicators), wealth (four indicators), rural indicator, altitude quintile indicators, and ethnicity indicators. Program controls include controls for the CB-NCP and CB-IMNCI health programs. The place of delivery is predicted using the linear probability shown in Appendix Table A.2. All specifications are estimated with district and month of birth fixed effects. Bootstrapped standard errors clustered at the district level in parentheses. Asterisks indicate significance at the following levels * p<0.1, ** p<0.05, and *** p<0.01.

Table A.10: Regression results using a Logit specification: The effect of CHX-NCP on neonatal mortality. - Dependent variable: Mortality by $\leq$1m.

	Sample			
	All	All	P(home birth) <0.5	P(home birth) >0.5
	(1)	(2)	(3)	(4)
CHX	-0.021*	-0.006	0.008	-0.053**
	(0.012)	(0.013)	(0.023)	(0.024)
1[P(home birth)>0.5]		-0.006		
		(0.006)		
CHX × 1[P(home birth)>0.5]		-0.029**		
		(0.014)		
CHX + CHX × 1[P(home birth)>0.5]		-0.035**		
		(0.015)		
Observations	21,750	21,613	6,778	10,846
Clusters	73	73	73	73
Control mean of dep. var	0.042	0.042	0.033	0.050
P-val (dif across sample)				0.048

Notes: Marginal effects. All specifications are estimated as logit models with the full set of demographic, SES, and program controls. Demographic controls include birth order (three indicators), five year maternal age group indicators, and gender. SES controls include education (three indicators), wealth (four indicators), rural indicator, altitude quintile indicators, and ethnicity indicators. Program controls include controls for the CB-NCP and CB-IMNCI health programs. We predict the place of delivery using the logit the specification shown in Appendix Table A.2. All specifications are estimated with district and month of birth fixed effects. Bootstrapped standard errors clustered at the district level in parentheses. Asterisks indicate significance at the following levels * p<0.1, ** p<0.05, and *** p<0.01.

Table A.11: Heterogeneity - very young mothers. Dependent variable: Mortality by $\leq$1m.

	Sample P(home birth) <0.5	Sample P(home birth) >0.5
	(1)	(2)
Program: CHX	0.004	-0.027**
	(0.009)	(0.011)
Mother 15-16y	0.076***	0.061
	(0.026)	(0.042)
Mother 15-16y × CHX	-0.086***	-0.044
	(0.026)	(0.059)
Interaction+Level	-0.083***	-0.071
	(0.026)	(0.061)
Observations	10,860	12,605
Clusters	73	73
Mean of dep. var.	0.042	0.050

Notes: All specifications are estimated as linear probability models using OLS with the full set of demographic, SES, and program controls. Demographic controls include birth order (three indicators), five year maternal age group indicators, and gender. SES controls include education (three indicators), wealth (four indicators), rural indicator, altitude quintile indicators, and ethnicity indicators. Program controls include controls for the CB-NCP and CB-IMNCI health programs. The place of delivery is predicted using the linear probability shown in Appendix Table A.2. All specifications are estimated with district and month of birth fixed effects. Bootstrapped standard errors clustered at the district level in parentheses. Asterisks indicate significance at the following levels * p<0.1, ** p<0.05, and *** p<0.01.

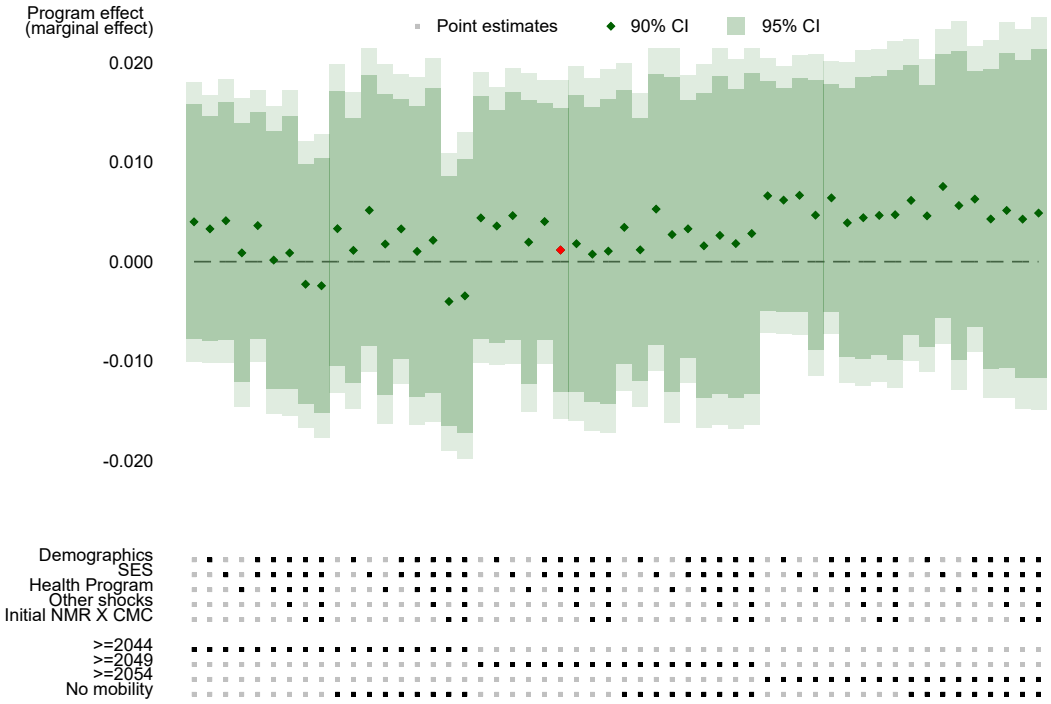
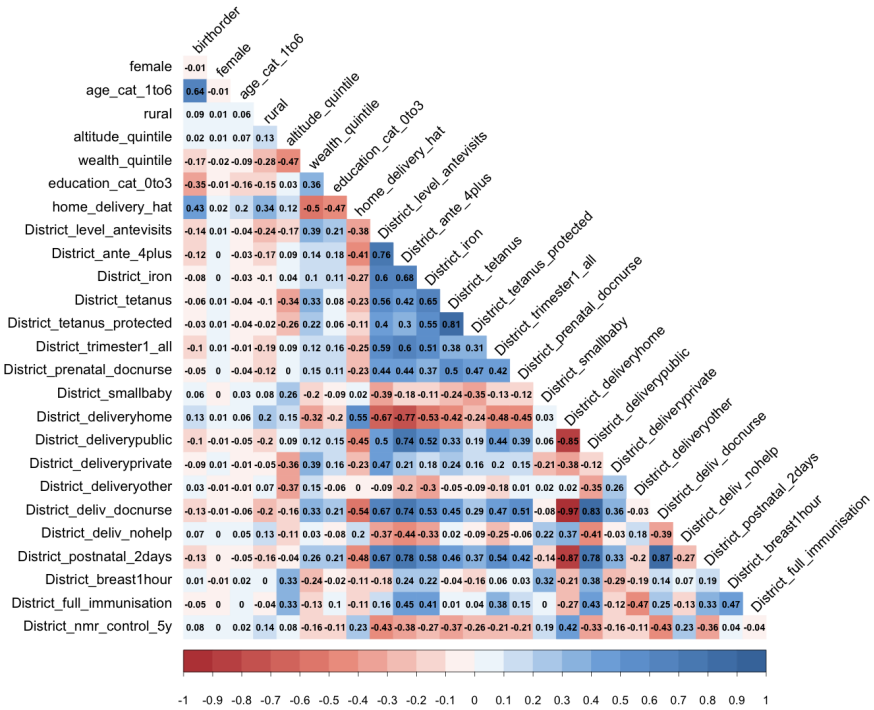


Figure A.2: Specification Curve for the sample P(home birth)<0.5
Notes: This chart shows estimates from running 128 different specifications defined by the combination of markers bellow the chart. Demographic controls include birth order (three indicators), five year maternal age group indicators, and gender. SES controls include education (three indicators), wealth (four indicators), rural indicator, altitude quintile indicators, and a Dalit ethnicity indicator. Program controls include controls for the CB-NCP and CB-IMNCI health programs. Other shocks refer to the earthquake on 25 April 2015, the Community Action for Nutrition Project, an Integrated Nutrition Program, and the Safe Delivery Incentive Program. Initial NMR $\times$ CMC is the initial neonatal mortality times a quadratic time trend. The place of delivery is predicted using the linear probability shown in Appendix Table A.2. The confidence intervals are based on 200 bootstrap iterations clustered at the district level.

Figure A.3: Correlation matrix - features included in the optimal policy



Notes: All district variables are averages taken over the five years preceding the survey. The names of variables defined at the individual level are self-explanatory. The "hat" in "home_delivery_hat" refers to an estimated probability of home delivery above 0.5. District variables are defined as follows: "District_level_antevisits" is the district average number of antenatal visits, "District_level_ante_4plus' is the district share of pregnancies with four or more antenatal visits, "District_iron" is the district share of pregnancies for which the mother received iron tablets during pregnancy, "District_tetanus" is the district share of pregnancies for which the mother received tetanus immunization during pregnancy, "District_tetanus_protected" is the district share of pregnancies for which the mother was protected against tetanus during pregnancy either because she received a tetanus injection during this pregnancy or due to prior immunization, "District_trimester1_all" is the district share of pregnancies for which the mother received an antenatal care visit during the first trimester, "District_prenatal_docnurse" is the district share of pregnancies for which the mother received antenatal care from a doctor or nurse, "District_smallbaby" is the district share of newborns whose mother reports as being smaller than average or very small at birth, "District_deliveryhome" is the district share of pregnancies delivered at home, "District_deliverypublic" ("District_deliveryprivate") is the district share of pregnancies delivered in a public (private) facility, "District_deliveryother" is the district share of pregnancies delivered neither in the home or in a facility, "District_deliv_docnurse" ("District_deliverynohelp") is the district share of pregnancies delivered with the help of a doctor or nurse (no help at all), "District_postnatal_2days" is the district share of newborns who received a postnatal check within two days of birth, "District_breast1hour" is the district share of newborns breastfed within an hour of birth, "District_full_immunisation" is the district share of children aged 12 to 60 months who have received a full course of immunization (1 BCG, 3 DPT, 3 Polio, and 1 Measles vaccines).

B Details of the Machine Learning Procedure

B. 1 Training the Causal Forest

To assess treatment effect heterogeneity we train a causal forest using the *grf* package in R (Athey et al., 2019; Tibshirani et al., 2021). Concretely, we proceed in the following two steps.

Step 1 We use regression forests to estimate the following two conditional mean functions

$$\mu_W = E[W|X = x] \tag{2}$$

$$\mu_Y = E[Y|X = x] \tag{3}$$

where W is equal to 1 if the child was born in a district and month where the CHX program was implemented and 0 otherwise, Y is 1 if the child died within the first month after birth and 0 otherwise, and X is a set of indicator variables capturing the district of birth, the month times year of birth, whether the CB-IMNCI program is implemented, and whether CB-NCP is implemented in the district. Using the fitted conditional mean functions we construct the residuals, $W - \mu_W$ and $Y - \mu_Y$.

Step 2 We use the residuals from the first step to train a causal forest which we use to estimate the conditional average treatment effects (CATEs):

$$\tau(X) = E[Y(1) - Y(0)|X = x] \tag{4}$$

where $Y()$ are the potential outcomes and X contains birth order, maternal education, maternal age, wealth, district, altitude, rural, predicted place of delivery, health programs, district, ethnicity; and district-level averages for: antenatal care (ANC) visits (timing and number), whether iron tablets were received during ANC visits, tetanus protection, place of delivery, postnatal visits, immunization rate, neonatal mortality, nurse or doctor-assisted delivery, and whether the baby was considered small at birth.

For the categorical variables (ethnicity and district) we use the sufficient representation approach where we compute and include group means of

the non-categorical variables based on the groups defined by the categorical variables.

In training the causal forest we tune all parameters by cross-validation. For non-tuned parameters we use the default settings, except that we set the forest to be clustered at the district level and we allow clusters to have different weights. The latter setting has very little practical implication in our setting. The chosen parameter settings are listed in Table B.1.

Table B.1: Causal Forest Settings

Setting	Value	Selection criteria
Number of trees	2000	Default
Clustering	District	Choice
Fraction of sample used to grow each tree	0.5	Cross-validation
Number of variables considered for each split	30	Cross-validation
Minimum size of a leaf node	5	Cross-validation
Fraction of sample used for splitting	0.5	Cross-validation
Prune empty leaves	True	Cross-validation
Maximum imbalance of a split (alpha)	0.05	Cross-validation
Penalization of imbalance splits	0	Cross-validation

Note: The table shows the parameter settings for the main causal forest. None of the parameters selected by cross-validation are different to the default setting.

Having specified the parameter settings, we grow a tree as follows:

(i) We sample 50% of the original analysis sample and 30 of the variables.

(ii) The sample selected in (i) is split into two equally sized sub-samples. One sub-sample is used to find the splitting structure of the tree. The second sample is used for populating the trees.

(iii) The sample for splitting found in (ii) is split into two groups (nodes) using the variable among the 30 selected in (i) that creates the best split. The best split maximizes treatment effect heterogeneity across the two groups and minimizes the variance in treatment effect heterogeneity within the groups.

(iv) The tree is grown by repeating step (iii) on the created groups until there is no valid split (for example if the number of observations is

smaller than 5) or if there is no split that improves the fit sufficiently. A group that is not split further is called a leaf.

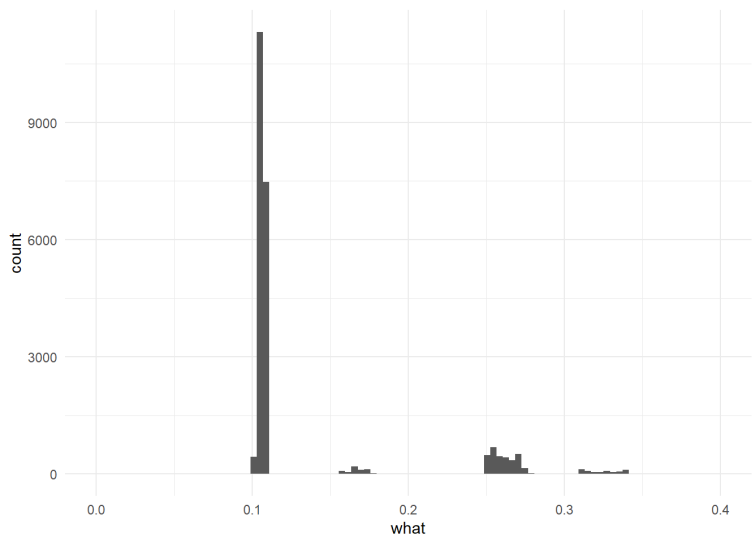
(v) Using the splitting structure found in (iv) the tree is populated using the second sub-sample created in (ii) and the outcomes are predicted based on these observations. In other words the hold out sub-sample for populating the trees runs through the decision tree (the splitting) and these observations are then used to obtain an estimate of the leaves' treatment effects.

Steps (i) to (v) creates a tree and this step is repeated 2000 times to create the forest. Having created the forest, an observation's predicted conditional treatment effect (CATE) is created based on the average predicted outcome for the leaves the observation ends up in across all trees where this observation was not used to split and populate the trees, i.e., based on the out-of-bag prediction.

B. 2 Distribution of propensity scores and covariates

Figure B.1 shows the distribution of propensity scores from the first step described above. These scores should be between 0 and 1 (not including 0 and 1), which is the case in our setting.

Figure B.1: Propensity scores for causal forest



Another important condition for the causal forest is that the features have common support across treatment status. Figures B.2 to B.4 suggest that this is the case in our setting.

Figure B.2: Propensity score weighted common support across for covariates

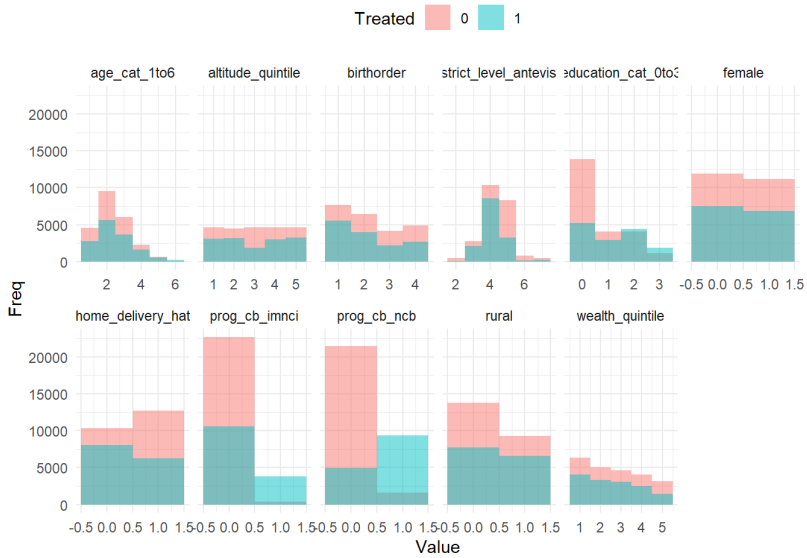


Figure B.3: Propensity score weighted common support across for ethnicity demeaned covariates

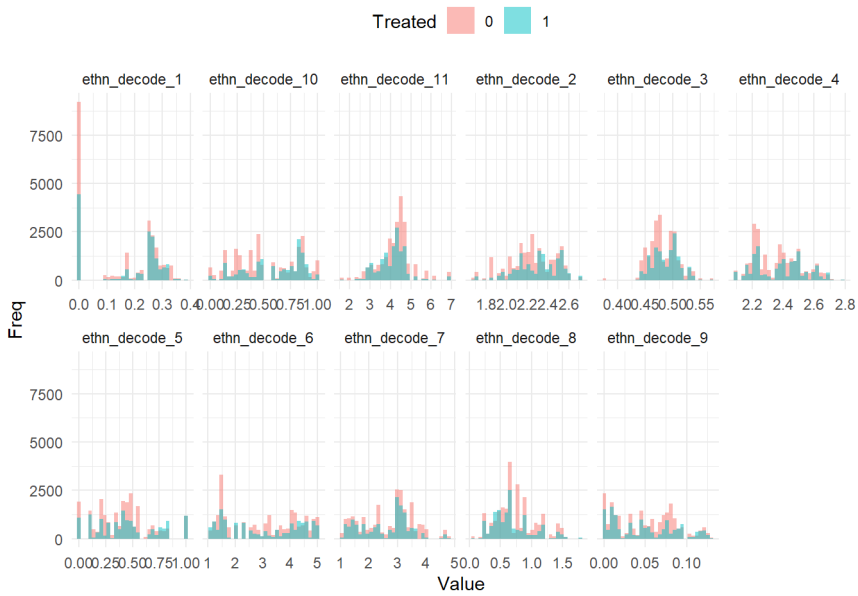


Figure B.4: Propensity score weighted common support across for district demeaned covariates

B. 3 Diagnostic test for forest fit

Table 5 in the main text presents results from a diagnostic test of the forest fit. The test is based on regressing the outcome on the mean (held-out) forest

prediction (A) and the (held-out) differential prediction on (B). If the forest captures the mean effect perfectly the coefficient on (A) should be 1, and if the forest captures heterogeneity perfectly, the coefficient on (B) should be 1. Moreover, if the coefficient on (B) is significantly greater than 0 it suggests treatment effect heterogeneity.

$$Y_i - \hat{m}^{-i}(X_i) = \alpha \overbrace{\bar{\tau}\left(W_i - \hat{e}^{-i}(X_i)\right)}^{(A)} + \beta \overbrace{\left(\hat{\tau}^{-i}(X_i) - \bar{\tau}\right)\left(W_i - \hat{e}^{-i}(X_i)\right)}^{(B)} + \varepsilon \tag{5}$$

B. 4 Alternative causal forest specifications

In Table B.2 we show results for four different specifications of the causal forest. Column (1) shows the main forest using the settings described above. Column (2) shows the results from training a forest using a smaller set of variables in step 2. This specification is used to obtain predictions of the CATEs for the five RCT samples. Column (3) shows the result of a specification based on the same variables used in specification (2), but where the wealth measure is adjusted to be comparable across RCT samples. In column (4) we show results of fitting a forest that mimics the specification in column (1) except that we use a continuous wealth measure.

Figure B.5 show the distribution of CATEs based on the specification shown in column (1) in table B.2.

Figure B.5: Distribution of Conditional Average Treatment Effects (CATEs)

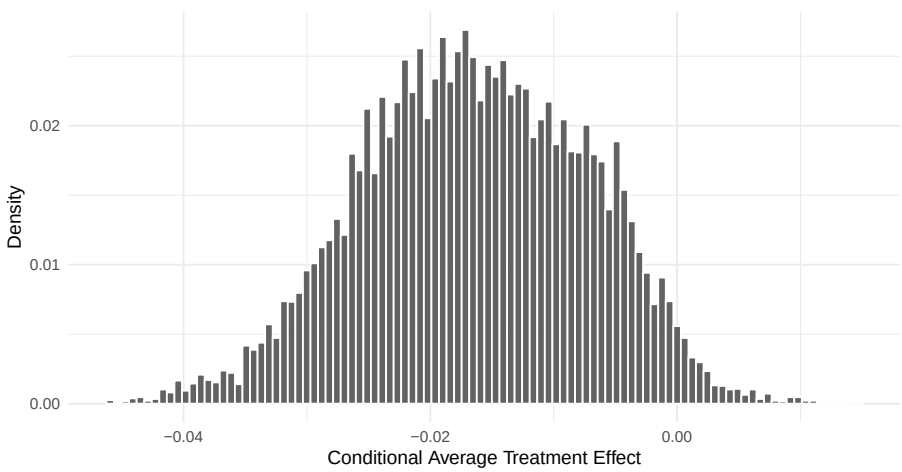


Table B.2: Various Causal Forest Specifications - Diagnostic Test, Average Treatment Effects, and Optimal Policies

	(1)	(2)	(3)	(4)
A. Omnibus diagnostic test for forest fit				
Mean Forest Prediction	1.215***	1.115***	1.116***	1.201***
	(0.273)	(0.265)	(0.229)	(0.271)
Differential Forest Prediction	0.806*	0.319	0.483*	0.967*
	(0.525)	(0.422)	(0.352)	(0.600)
B. Doubly Robust Average Treatment Effects				
Full sample	-0.020***	-0.020***	-0.021***	-0.017***
	(0.003)	(0.003)	(0.003)	(0.004)
Predicted home delivery	-0.030***	-0.031***	-0.032***	-0.029***
	(0.004)	(0.004)	(0.003)	(0.006)
Predicted facility delivery	-0.007*	-0.008*	-0.007	-0.005
	(0.004)	(0.004)	(0.004)	(0.005)
C. Optimal policies - reduction in NMR				
All variables	-0.022***	-0.022***	-0.023***	-0.022***
	(0.002)	(0.002)	(0.002)	(0.002)
	[0.860]	[0.860]	[0.842]	[0.860]
Reduced set of variables	-0.022***	-0.022***	-0.023***	-0.022***
	(0.002)	(0.002)	(0.002)	(0.002)
	[0.840]	[0.840]	[0.842]	[0.840]
District variables only	-0.022***	-0.022***	-0.022***	-0.022***
	(0.002)	(0.002)	(0.002)	(0.002)
	[0.823]	[0.823]	[0.840]	[0.823]
Causal forest variables	Full	Reduced	Reduced	Full
Wealth measure	DHS	DHS	Comparable	Continuous
Orthogonalisation	RF	RF	RF	RF

Notes: Column (1) shows the results for the main specification. Column (2) shows results for the specification used to compare predicted CATEs across samples. Column (3) is as in column (2) but based on a wealth measure we construct to make it comparable across specifications. Column (4) shows results using a continuous wealth measure instead of quintiles. Column (5) shows results using OLS to orthogonalise treatment and outcomes. Standard errors clustered at the district level in parentheses. The share treated int he optimal policies are reported in squared brackets. Asterisks indicate significance at the following levels * p<0.1, ** p<0.05, and *** p<0.01.