Natalie Carvalho
Slawa Rokicki

The impact of India's JSY conditional cash transfer programme

A replication study

June 2015

Replication Paper 6

Health

International Initiative for Impact Evaluation

About 3ie

3ie is an international grant-making NGO promoting evidence-informed development policies and programmes. We are the global leader in funding and producing high-quality evidence of what works, how, why and at what cost. We believe that better and policy-relevant evidence will make development more effective and improve people's lives.

3ie Replication Paper Series

The 3ie Replication Paper Series is designed to be a publication and dissemination outlet for internal replication studies of development impact evaluations. Internal replication studies are those that reanalyse the data from an original paper in order to validate the results. The series seeks to publish replication studies with findings that reinforce an original paper, as well as those that challenge the results of an original paper. To be eligible for submission, a replication study needs to be of a paper in 3ie's online Impact Evaluation Repository and needs to include a pure replication. 3ie invites formal replies from the original authors. These are published on the 3ie website together with the replication study.

The **3ie Replication Programme** also includes grant-making windows to fund replication studies of papers identified on the candidate studies list. Requests for proposals are issued once or twice a year. The candidate studies list includes published studies that are considered influential, innovative or counterintuitive. The list is updated periodically based on 3ie staff input and outside suggestions. The aim of the 3ie Replication Programme is to improve the quality of evidence from development impact evaluations for use in policymaking and programme design.

About this report

This paper was funded through 3ie's Replication Window with generous funding from an anonymous donor. All content is the sole responsibility of the authors and does not represent the opinions of 3ie, its donors or the 3ie Board of Commissioners. Any errors and omissions are the sole responsibility of the authors. Comments and queries should be directed to the corresponding author, Slawa Rokicki, srokicki@fas.harvard.edu.

Suggested citation: Carvalho, N and Rokicki, S, 2015. *The impact of India's JSY conditional cash transfer programme: A replication study. 3ie Replication Paper 6*. Washington, DC: International Initiative for Impact Evaluation (3ie).

3ie Replication Paper Series executive editor: Annette N Brown
Managing editor: Benjamin DK Wood
Assistant managing editor: Jennifer Ludwig
Production manager: Jennifer Ludwig
Layout assistants: Kara Ingraham, Jane Burke
Copy editor: Jaime L Jarvis
Cover design: John F McGill
Proof reader: Yvette Charboneau
Printer: VIA Interactive

The impact of India's JSY conditional cash transfer programme: A replication study

Natalie Carvalho
University of Melbourne, School of Population and Global Health

Slawa Rokicki
Harvard University, Department of Health Policy, Evaluative Sciences and Statistics

3ie Replication Paper 6
June 2015

International Initiative for Impact Evaluation

Acknowledgments

The authors would like to acknowledge 3ie for funding, technical review and support throughout the study. We would also like to thank the reviewers of our proposal and final replication report. Finally, we are grateful to the authors of the Lim *et al.* (2010) paper, who provided us with their code and were helpful in answering our questions.

Abstract

Conditional cash transfer programmes are becoming increasingly popular in low- and middle-income countries, with the goal of improving access to health and social services and reducing inequities in access and outcomes for the poor and marginalised. India's conditional cash transfer programme, *Janani Suraksha Yojana* (JSY), established in 2005, is one of the largest such programmes in the world. Along with small payments to community health workers, it provides financial incentives to pregnant women to encourage them to deliver in health facilities. Lim *et al.*'s *Lancet* article, 'India's Janani Suraksha Yojana, a conditional cash transfer programme to increase births in health facilities: an impact evaluation' (2010), was the first formal statistical impact evaluation of the programme across the whole of India.

This replication study, through robustness checks and additional model specifications, re-examines this recent work on the effect of financial incentives for women through the programme on reproductive health coverage indicators and perinatal and neonatal mortality. Of three analytic approaches taken by Lim *et al.* (2010), this replication focused on exact matching analysis, using data from round three of India's district-level household survey (DLHS-3).

We found the original authors' results to be replicable and robust to various changes in model specifications and analysis. We were able to replicate quite closely the national and subnational results reported by Lim *et al.*

We conducted several additional analyses as robustness checks including alternative matching estimators, analyses to account for differential implementation of the programme and random effects models to examine state-level heterogeneity in results. We found meaningful heterogeneity across states and districts in the effects of JSY on reproductive health coverage indicators and mortality outcomes.

Accounting for state- and district-level heterogeneity has important implications for understanding the effectiveness of this programme.

Keywords: replication study, conditional cash transfers, maternal health, neonatal mortality, skilled birth attendance

Contents

List of figures and tables

Appendix tables

Appendix figures

Abbreviations and acronyms

ANC	Antenatal care
ASHA	Accredited social health activist
BPL	Below poverty line
CI	Confidence interval
DLHS	District-level household survey
JSY	*Janani Suraksha Yojana*
NRHM	National Rural Health Mission
OR	Odds ratios

1. Introduction

India's conditional cash transfer programme, *Janani Suraksha Yojana* (JSY), is one of the largest programs of its kind in the world (Lim *et al.* 2010). Launched under the National Rural Health Mission, JSY provides financial incentives to pregnant women to encourage them to deliver in health facilities. Performance-based cash payments are also offered to community health workers, called 'accredited social health activists (ASHAs)', to facilitate in-facility deliveries and promote other positive reproductive and child health behaviours (IIPS 2010). Eligibility and financial incentives vary across states, with priority given to women in 10 low-performing states. In these 'high focus' states, all women are eligible for the programme and the cash incentives are higher than in the other states. In other states, women are eligible for the cash benefit only for their first two live births, and only if they have a government-issued Below Poverty Line (BPL) card or if they are from a scheduled (low) caste or tribe (MHFW 2006). The JSY programme aims to increase access to safe pregnancy and delivery services, with the overall goal of reducing maternal and neonatal mortality and morbidity.

Despite substantial progress made over the last decade in reducing maternal and under-5 mortality, India still accounts for the world's highest number of maternal, neonatal and child deaths, with persistently high mortality rates among these groups (Lozano *et al.* 2011; Black *et al.* 2010; Bhutta *et al.* 2010). Globally, mortality during the neonatal period has experienced slower reductions (2 per cent per year since 1990) compared with reductions in maternal mortality (2.6 per cent) and child mortality (3.4 per cent), with even slower progress achieved in reducing stillbirths (Bhutta *et al.* 2014). Facility-based care, in particular through interventions during labour and delivery, are likely to make the biggest impact in reducing maternal deaths, neonatal deaths and stillbirths (Bhutta *et al.* 2014).

The annual number of JSY beneficiaries grew rapidly, from 739,000 per year in 2005–2006 to more than 11 million in 2010-2011 (MHFW 2011). The programme reflects an important component of the Indian government's spending on health. With the launch of the National Rural Health Mission (NRHM), public spending on health increased by nearly 2.6 times between the 2004–2005 and 2009–2010 fiscal years (MHFW 2010), with a budget allocation for JSY estimated at US$342 million in the 2009–2010 fiscal year (Lim *et al.* 2010).

2. Motivation and literature review

2.1 Overview of conditional cash transfer programmes

Conditional cash transfer programmes are being increasingly introduced in low- and middle-income countries, with the goals of improving access to health and social services and related outcomes and reducing inequities in access and outcomes for the poor and marginalised. A Cochrane review found evidence that conditional cash transfer programmes had a positive impact on the use of health services and the uptake of preventive services by children and pregnant women (Lagarde *et al.* 2009). A recent systematic review on the effects of conditional cash transfer programmes on maternal and new-born health reported that such programmes have increased antenatal visits, skilled attendance at birth, and delivery at a health facility and have reduced the incidence of low birth weight babies (Glassman *et al.* 2013). Although both of these reviews include evidence from some programmes that have been evaluated using well-designed studies, more rigorous evaluation is needed to assess the impact of these programmes on health and utilisation outcomes in different settings (Lagarde *et al.* 2009; Glassman *et al.* 2013).

2.2 Assessments of India's JSY programme and the Lim *et al.* evaluation

As one of the largest cash transfer programmes in the world, JSY has received much attention since its rollout in 2005. Evaluating JSY is crucial to understanding its effectiveness in improving maternal and neonatal health outcomes and reducing existing inequities in access and outcomes. Understanding the programme's effectiveness is important not only for policymakers in India, but also in offering lessons for other countries with low rates of institutional delivery and poor reproductive health outcomes.

Lim *et al.* (2010) conducted the first formal statistical impact evaluation of JSY across the whole of India. Prior assessments had been more descriptive in nature (Devadasan *et al.* 2008) or geographically limited (UNFPA 2009; Sharma 2009), or had considered very limited outcomes (Satapathy 2009). Lim *et al.* (2010) employed three analytic approaches (exact matching with logistic regression, with-versus-without, and district-level differences-in-differences) to estimate the effect of maternal receipt of financial incentives from JSY on levels of institutional delivery, skilled birth attendance, antenatal care, and maternal, perinatal and neonatal mortality. They found significant positive effects on antenatal care, institutional delivery and skilled birth attendance. In two of their three analytic approaches, they also found a reduction in perinatal and neonatal deaths. The authors were unable to detect an effect of the programme on maternal deaths. The Lim *et al.* (2010) paper has been an influential study, widely cited in the literature and discussed in international health economics and maternal health conferences over the last few years (World Bank 2012, Morgan *et al.* 2011).

Despite the study's influence, criticisms have been expressed regarding Lim *et al.*'s findings. Recent research has cited problems with the JSY programme, including slow or uneven implementation, corruption, poor quality of care, out-of-pocket costs of in-facility delivery and additional challenges faced by women who take part in JSY, leading to claims that the programme has not been as successful as the results seem to suggest (Sukla 2012; Das *et al.* 2011; Mazumdar, Mills and Powell-Jackson 2011). Das *et al.* (2011) called for a further review of JSY, pointing out two main methodological limitations. First, they suggest that there is ambiguity in the survey question asking women about enrolment in JSY, given that the question includes state-specific schemes. Second, they note that although JSY was established in 2005, it took some time to become operational. In their response, Lim *et al.* (2011) disagree that reference to state-specific schemes makes the question about participation ambiguous. They point out that this wording was necessary because JSY could have different names in different states and local languages. Regarding the length of time between

JSY's establishment and implementation, Lim *et al.* (2011) attempt to minimise the effect of capturing earlier schemes by restricting the analysis to births in the 12 months preceding the survey. Although we believe this is a reasonable solution, as we will demonstrate in the following sections, we do not believe this restriction was carried out on all analyses presented in their paper, particularly the final analysis estimating the effect of JSY.

Mazumdar, Mills and Powell-Jackson (2011) conducted the second national formal statistical assessment of JSY using the same data as Lim *et al.* (2010) but taking a different statistical approach. They carry out a difference-in-differences analysis to estimate the impact of JSY using an instrumental variable approach, with an indicator for when JSY was introduced in each district. Compared with Lim *et al.*'s (2010) results, Mazumdar, Mills and Powell-Jackson (2011) found a significant but smaller impact of JSY on in-facility delivery, little to no impact on antenatal care, and no effect on neonatal or early neonatal mortality. Their approach may provide improved estimates; it may allow for improved control of district-level unobservables such as management and capacity of district health authorities, which can generate spurious relationships between JSY and the outcomes. However, instrumental variable also increases standard errors due to uncertainty in the first stage, which may be why Mazumdar, Mills and Powell-Jackson (2011) find a coefficient on neonatal mortality outcomes that is of the same sign and of comparable size to Lim *et al.* (2010), –1.5 compared to –2.3 per 1,000 live births, but with larger confidence intervals (CIs).

It is worth noting that Lim *et al.* (2010) also carry out a district-level differences-in-differences approach, which similarly enables them to control for district-level unobservable variables. They find no significant effects of JSY on perinatal and neonatal mortality outcomes through this approach, although their point estimates were much larger, as were the CIs. However, they find quite similar treatment effect sizes for the reproductive health coverage outcomes (institutional delivery, skilled birth attendance and antenatal care), compared with their other two approaches. Others have tried to evaluate the effect of JSY on the maternal mortality ratio (Randive, Diwan and De Costa 2013), but no study to date has found any effect. In part, the data currently available are unlikely to be sufficiently powered to detect any meaningful changes in maternal mortality.

2.3 Overview of the replication study

Coverage of institutional delivery, skilled birth attendance, antenatal care and perinatal and neonatal mortality are key outcomes for understanding the success or failure of the programme. Replicating the results of the Lim *et al.* (2010) study in light of these outcomes is important to confirm the validity and robustness of the results and potentially address some of the criticisms.

This replication study begins with a pure replication, followed by a measurement and estimation analysis consisting of two main components. The first measurement and estimation analysis tests the robustness of Lim *et al.*'s matching algorithm, along with further restrictions of the data to try to reduce biases associated with differential implementation of the scheme across states and districts. The second uses multilevel modelling methods to explore state-level heterogeneity. We believe each of these components can contribute to a better understanding of JSY's impact. To our knowledge, no additional data exists beyond what Lim *et al.* (2010) and Mazumdar, Mills and Powell-Jackson (2011) used to evaluate the programme. Although the fourth wave of the DLHS (2012–2013) was completed, it is not yet available to the public. We are thus limited to the same data, which covers the period just before the start of JSY and the programme's first few years of implementation. Furthermore, the use of observational, cross-sectional data limits our ability to address an important issue regarding reverse causality between JSY and the institutional delivery outcome, because women receive the cash incentive when they deliver in facilities (Mazumdar, Mills and Powell-Jackson 2011). To explore this further, we carried out additional analyses to identify some of the potential causal pathways through which the programme might be operating.

The analyses in the pure replication and the measurement and estimation analysis were pre-specified in a replication plan posted on the 3ie Replication website (Carvalho and Rokicki 2013); a link to the plan is provided in the list of references.

2.4 Main replication questions

This study replicates three sets of results from the Lim *et al.* (2010) study:

1. *Participation*: What are the characteristics of JSY beneficiaries? Is the programme reaching the target population?

2. *Impact on coverage*: What are reasonable estimates of the programme's impact on the following reproductive health coverage indicators: antenatal care, institutional delivery and skilled birth attendance, nationally, separately for high-focus and non–high-focus states and across selected key states?

3. *Impact on health*: Finally, and perhaps most relevant for policy, what is JSY's impact on health outcomes? Although Lim *et al.* (2010) found no impact on maternal mortality, their findings of small but significant reductions in neonatal and perinatal mortality in two of their three methodological approaches are important, have been more controversial and would be most interesting to replicate and validate.

Mazumdar, Mills and Powell-Jackson (2011) have already investigated variants in delivery location and skilled birth attendance outcomes, such as the type of facility chosen for delivery, the health provider in attendance and type of procedure(s) performed. We have not repeated these analyses.

3. Pure replication

This replication begins by validating the original results of Lim *et al.* (2010) – hereafter also referred to as 'the authors'. We first replicate the summary statistics of JSY uptake and the logistic regression to assess the associations between maternal receipt of financial assistance from JSY and individual and household characteristics. The authors' most controversial findings were the effect sizes of the estimates on antenatal care, institutional delivery and neonatal and perinatal health outcomes (Mazumdar, Mills and Powell-Jackson 2011). Because the matching analysis resulted in the most conservative yet statistically significant estimated treatment effect for these four outcomes, we focus on the exact matching analytical approach and replicate the results for these findings. We do not replicate the authors' two other analytic approaches. This is because the estimated treatment effects from the with-versus-without analysis were not statistically different from the exact matching analysis, and, similar to the findings of Mazumdar and colleagues (2011), Lim *et al.* (2010) found no significant effect of JSY on perinatal and neonatal deaths in the district-level differences-in-differences analysis. Although reducing maternal mortality was a primary aim of the programme, the authors were unable to detect a significant effect of JSY on maternal mortality through their differences-in-differences analysis. The CIs around the estimated treatment effect were very wide, giving little meaning to the computed effect, and the authors speculate that the survey was unpowered to detect the effect of JSY on the number of maternal deaths (Lim *et al.* 2010).

We maintained all of the authors' original assumptions and methods for aggregating districts, estimating household wealth and characterizing categorical variables, and implemented the same exact-matching analysis with logistic regression. During the process of the pure replication, we independently constructed the variables from the raw data and re-estimated the results using the study's methodologies. The following section documents any resulting discrepancies we found in variable construction and analysis methods.

3.1 Data and methods

The data came from round three of India's district-level household survey (DLHS-3), which sampled 720,320 households between late 2007 and early 2009. Districts with low coverage of interventions for maternal and child health were oversampled.

Lim *et al.* (2010) used multivariate logistic regression to evaluate the association between receipt of financial assistance from JSY by women for their most recent birth and individual characteristics. For the analysis on the effect of JSY on reproductive health coverage indicators and health outcomes, the authors used coarsened exact matching to preprocess the data in an effort to make the treated group (those who received JSY financial assistance) as similar to the control group as possible. The authors matched on state of residence, urban or rural residence, BPL card ownership, wealth quintile, caste, education, parity and maternal age. They then used logistic regression to estimate the effect of JSY on three reproductive health coverage indicators – the likelihood that a woman attended at least three antenatal care visits, the likelihood that a woman gave birth in a health facility and the likelihood that a woman had skilled birth attendance (defined as birth in a health facility or with a skilled attendant present outside of a health facility) – and two health outcomes – perinatal death (stillbirth after 28 weeks of pregnancy or death of a child within first week after being born alive) and neonatal death (death of a child within the first month after being born alive). The authors use a number of covariates in the logistic regression to provide additional control for confounding.

The authors' analysis uses Stata and R. After the working dataset is created in Stata, it is exported to R (version 3.1.0) and uses the MatchIt package (version 2.4-13) to perform coarsened exact matching. It is then imported back to Stata for logistic modelling. The authors provided us with nearly all their Stata and R code, and we conducted the analysis in the same way they did, using the same version of MatchIt.

3.2 Creation of working dataset

We identified a small number of very minor coding issues and one more important coding error in the Stata do-files and R code. First, the authors' programming code separates neonatal mortality into early and late neonatal deaths. The first minor coding issue we identified was a simple mistake defining early neonatal mortality. We believe the authors corrected this coding mistake in a later do-file; once we made the appropriate modification we were able to replicate nearly identically the national-level mean coverage and mortality outcomes reported in the authors' Table 2. The second very minor coding issue we identified was in the generation of the *number of births* variable. This variable was coded in such a way that it missed 16 women who had births after 2004.

Third, we identified a coding issue that, although also minor, essentially renders the working dataset irreproducible. This arose in the creation of a *birth index* variable, which is used to order births for each woman. The birth index is created after sorting the data in a non-unique manner that includes ties, occurring when a woman has more than one birth recorded in the same year. This occurs nearly 3,500 times in the data, of which the majority (95 per cent) are multiple births. However, when Stata is asked to sort among non-unique observations, it chooses to sort the ties randomly, rendering the dataset irreproducible.

We identified one more important and unresolved coding issue in the creation of the *date of last birth* variable. The authors' code appears to misalign birth month and birth year when creating this variable, and we believe the result was 26,585 observations (13 per cent) incorrectly dropped from the analysis. Although this is a large number of dropped observations, they were essentially dropped at random, so correcting for this did not appear to have a major effect on the results apart from increasing the sample size. Appendix Table 1 provides details on all of the coding issues we identified.

Table 1 below compares our sample sizes and those presented by the authors in their Table 1 (logistic regression of the association between receipt of JSY and individual characteristics) and their Web Table 1 (logistic regression results from the exact matching analysis). We show the authors' sample sizes (row 1), sample sizes from our attempt to exactly match their analysis using the code as given to us without any modifications (row 2), sample sizes when we resolved the coding error for early neonatal mortality (row 3) and sample sizes when we resolved all of the coding issues previously identified (row 4).

We carried on with our pure replication efforts using two working datasets. The first (Working Dataset 1) contains the adjusted definition for the *early neonatal mortality* variable, as we are quite certain the authors had made this adjustment in a later do-file. For the second dataset (Final Working Dataset), we resolved all the coding issues identified above.

Table 1: Number of observations (N) for replication efforts compared to Lim *et al.* (2010)

	Logistic regression of beneficiaries of JSY	Exact matching analysis (restricted to births after 1 January 2004)
Lim *et al.* N	182,869[a]	158,907[b]
Replication N using authors' code as provided	182,217	162,391
Replication N using Working Dataset 1	182,764	163,819[c]
Replication N using Final Working Dataset	208,474	189,533[c]

Note: [a] Lim *et al.* (2010) Table 1, column 1 (national)
[b] Lim *et al.* (2010) Web Table 1, column 2 (in-facility births outcome)
[c] Matching run separately for each outcome (explained in section 3.7)

We were not provided with the full code required to replicate the authors' household wealth index, although the authors did provide the final wealth index for each household, which we were able to use to replicate their results. Instead, we used the wealth variable available from the survey and created our own wealth index using factor analysis, based on the same household assets used by the authors that were available in both rounds of the DLHS: type of toilet, type of house and type of cooking fuel; source of water; and ownership of a fan, television, motorcycle, car and telephone. Results were similar regardless of which wealth index was used to classify households into wealth quintiles and deciles.

3.3 National-level means

We compared our estimates for national level means of the reproductive health coverage indicators and health outcomes, with the authors' estimates reported in their Table 2 (Table 2, below). These mean values were calculated using data from the DLHS-3 for births in the 12 months prior to the survey. With the exception of the point estimate for perinatal mortality and very minor differences in the computed 95 per cent CIs, our replicated values match exactly with the mean values calculated by the authors. This first validation step was useful in determining that our efforts to code the main outcome variables of interest matched the authors' code.

Table 2: National-level means

	Lim *et al.* (2010) Table 2 (exact matching) Mean (95% CI)	Pure replication using Working Dataset 1 Mean (95% CI)	Pure replication using Final Working Dataset Mean (95% CI)
Antenatal care, three visits	53.6% (53.0 to 54.3)	53.6% (53.2 to 54.1)	52.0% (51.6 to 52.4)
In-facility births	54.1% (53.5 to 54.8)	54.1% (53.7 to 54.6)	52.2% (51.7 to 52.6)
Skilled birth attendance	59.3% (58.7 to 60.0)	59.3% (58.8 to 59.7)	57.4% (57.0 to 57.8)
Perinatal deaths (per 1,000 pregnancies)	37.3 (35.6 to 39.0)	37.1 (35.4 to 38.7)	37.1 (35.6 to 38.6)
Neonatal deaths (per 1,000 live births)	30.3 (28.8 to 31.9)	30.3 (28.8 to 31.8)	30.6 (29.2 to 31.9)

Note: National-level means are percentage estimates (95 per cent CI) calculated from survey data for births in the past 12 months. Working Dataset 1 contains an adjusted definition for *early neonatal mortality*. Final Working Dataset resolves all coding issues.

3.4 Characteristics of JSY beneficiaries

We were able to replicate the authors' Table 1 ('Multivariable logistic regression of association between receipt of financial assistance from JSY by women for their most recent birth and individual characteristics') very closely using Working Dataset 1 (see Appendix Table 2). Rerunning this logistic regression on the Final Working Dataset resulted in very few differences in the odds ratios (OR), so these results are not presented. As the authors found, receipt of cash assistance from JSY was higher for younger mothers with middle levels of education and middle bands of wealth. Women in scheduled castes and tribes were more likely to have received financial assistance through the programme. There was wide variation in uptake of JSY across states, even across high-focus states where all women were eligible.

3.5 Replication of figures

We were able to closely replicate the authors' Figure 2 and Figure 3 (see Appendix Figures 2 and 3). Although both figure labels indicate that the data were subset to women who gave birth in the 12 months before the interview date, we believe the full dataset was used for Figure 3. We show in side-by-side bar graphs our replication efforts of the authors' Figure 3 when restricting the dataset to the previous 12 months and without restricting to 12 months (Appendix Figure 3). The unrestricted bars much more closely resemble the results from the authors' paper. We found no such discrepancy with Figure 2 and were able to reproduce it (see Appendix Figure 2).

3.6 Weights

In the exact matching analysis, the authors incorporate survey weights when analysing the marginal effects of the logistic regression. They do not account for the survey weights during the matching process or in the logistic regression. Although there is some inconsistency in the literature about the use of survey weights in matching analysis, we believe it is better to match on the survey weight (or on the variables that created the survey weight) in the matching algorithm (DuGoff *et al.* 2014; Gelman 2007).

The weights then used in the logistic analysis incorporate the matching weights and the survey weights. Although this is more intuitive in propensity score matching, it is possible to match on a coarsened survey weight in a coarsened exact matching analysis. We conducted the analysis incorporating the survey weights into the matching algorithm and did not find a meaningful difference in the final results (results not shown).

3.7 Dropping outcome variables

Another issue with the matching and logistic regression code as received from the authors is that all outcomes with missing values are dropped before conducting the matching and logistic regression. Although eliminating missing outcomes is necessary for matching to be carried out correctly, doing so for all outcomes at the same time results in all stillbirths dropping out of the analysis. This occurs because the neonatal death outcome is coded as missing for all stillbirths. Our solution was to rerun the analysis separately for the neonatal death outcome (with live births as the denominator), and again for all other outcomes (with pregnancies as the denominator).

Table 3 presents our efforts to replicate the estimated treatment effect (authors' Table 2, exact matching approach). Treatment effects are expressed as a change in predicted probability as a result of receipt of financial assistance from JSY. The first column shows the authors' results. The second column shows our pure replication using Working Dataset 1. The third column also uses Working Dataset 1 but uses our solution for addressing the issue of dropping all missing outcomes before the matching and logistic regression. Finally, the last column uses the Final Working Dataset along with our method for addressing the issue of dropping the missing outcomes. We found that using this method brought us much closer to the authors' final treatment effect results, particularly for mortality outcomes, leading us to believe that the authors may have addressed this issue in a later do-file.

Table 3: Estimated treatment effect at the national level

	Lim *et al.* (2010) Table 2 (exact matching) Mean (95% CI)	Pure replication using Working Dataset 1 Mean (95% CI)	Pure replication using Working Dataset 1, modification for dropping missing outcomes Mean (95% CI)	Pure replication using Final Working Dataset, modification for dropping missing outcomes Mean (95% CI)
Antenatal care, three visits	10.7% (9.1 to 12.3)	11.0% (10.0 to 11.9)	11.0% (10.0 to 11.9)	11.0% (10.1 to 11.8)
In-facility births	43.5% (42.5 to 44.6)	44.0% (43.2 to 44.8)	43.9% (43.1 to 44.7)	45.7% (45.0 to 46.4)
Skilled birth attendance	36.6% (35.6 to 37.7)	37.1% (36.3 to 37.9)	37.1% (36.3 to 37.9)	39.0% (38.3 to 39.7)
Perinatal deaths (per 1,000 pregnancies)	-3.7 (-5.2 to -2.2)	-2.1 (-3.2 to -1.0)	-3.6 (-5.0 to -2.3)	-3.8 (-5.2 to -2.5)
Neonatal deaths (per 1,000 live births)	-2.3 (-3.7 to -0.9)	-2.7 (-4.0 to -1.4)	-2.7 (-4.0 to -1.4)	-2.3 (-3.6 to -1.0)

Note: Treatment effects are expressed as the change in predicted probabilities. Working Dataset 1 contains the adjusted definition for early neonatal mortality. Final Working Dataset resolves all coding issues.

Making the coding modifications detailed in the appendix and above led to closer estimates compared to the authors' final results (Table 3), particularly for the mortality outcomes, however it also led to a larger sample size (Table 1). We believe the treatment effect for the health outcomes may have been run for neonatal mortality and/or perinatal mortality separately, but it is unclear based on the similar reported sample sizes for all outcomes in the authors' Appendix Table 1.

3.8 Subnational analyses

We repeated the pure replication for high-focus, north-east and other states analyses using the Final Working Dataset with our method for dropping the missing outcomes prior to running the analysis. Table 4 shows the results, which correspond to the authors' Table 3 results. Across all groups of states and all coverage and health outcomes, our results are very close to the authors' findings.

Table 4: Estimated treatment effect by high-focus, north-east and other states

	High-focus states Mean (95% CI)	North-east states Mean (95% CI)	Other states Mean (95% CI)
Lim et al. *(2010) Table 3 (Exact matching)*			
Antenatal care, three visits	11.1% (9.9 to 12.2)	14.1% (10.5 to 17.8)	3.2% (1.6 to 4.8)
In-facility births	63.8% (63.0 to 64.6)	36.0% (32.6 to 39.4)	6.6% (4.9 to 8.2)
Skilled birth attendance	58.7% (58.0 to 59.5)	31.6% (28.5 to 34.6)	4.9% (3.6 to 6.3)
Perinatal deaths (per 1,000 pregnancies)	-2.9 (-5.1 to -0.6)	-2.5 (-6.1 to 1.1)	-5.0 (-7.1 to -2.9)
Neonatal deaths (per 1,000 live births)	-1.4 (-3.5 to -0.7)	-0.1 (-3.2 to 3.0)	-4.2 (-6.1 to -2.2)
Pure replication using Final Working Dataset, with modification for dropping missing outcomes			
Antenatal care, three visits	10.9% (9.8 to 11.9)	14.6% (11.3 to 17.9)	3.7% (2.9 to 4.5)
In-facility births	65.0% (64.3 to 65.7)	35.6% (32.6 to 38.7)	7.6% (6.7 to 8.6)
Skilled birth attendance	60.1% (59.4 to 60.7)	31.2% (28.5 to 33.9)	6.0% (5.3 to 6.8)
Perinatal deaths (per 1,000 pregnancies)	-2.8 (-4.8 to -0.7)	-2.3 (-5.2 to 0.6)	-5.8 (-7.6 to -4.1)
Neonatal deaths (per 1,000 live births)	-1.1 (-3.1 to 1.0)	0.1 (-2.8 to 2.9)	-5.0 (-6.7 to -3.3)

Note: Treatment effects are expressed as the change in predicted probabilities. Final Working Dataset resolves all coding issues.

3.9 Pure replication conclusions

We were able to closely replicate the authors' national and subnational results as reported in Lim *et al.* (2010). Some of the coding issues we identified had most likely been corrected by the authors before they conducted the final analyses, but we are unable to determine which these were, because the Stata code files we received from the authors did not appear to be the final versions. However, based on our pure replication analysis, we find that the authors' results are robust.

4. Additional analyses

All analyses that follow were specified in a replication plan posted on the 3ie Replication website (Carvalho and Rokicki 2013); a link to the plan is provided in the list of references. If analyses deviate from the plan, a justification is provided. All analyses were carried out using the Final Working Dataset.

4.1 Alternative matching estimates

We begin by alternating the matching estimator. Lim *et al.* (2010) employ coarsened exact matching to preprocess the data and make the treatment variable as independent of background characteristics as possible. To test the robustness of their results, we implement propensity score matching, a more widely used matching technique, to compare the balance and robustness of the results under this matching method (Rosenbaum and Rubin 1983). Propensity score matching allows us to use a wider set of covariates than coarsened exact matching, without losing observations due to empty bins. Coarsened exact matching limits in the number of covariates that can be used due to sample size of the matching bins. Expanding the matching algorithm to include more covariates may reduce bias from confounding and may impact the results.

We conduct three algorithms for the propensity score model. In model 1, we use the same set of covariates as the original paper, in their coarsened forms, to compare the results from propensity score matching and the results from coarsened exact matching. We also maintain the authors' procedure for incorporation of survey weights. In model 2, we match on additional covariates – the full set of covariates used in the logistic regression analyses. We also employ our procedure for survey weights; that is, we match on survey weight and multiply the propensity weight by the survey weight in the logistic analysis. Third, we conduct the same model as model 2, but we include district fixed effects in the propensity score model and in the subsequent logistic regression.

All propensity score algorithms include a calliper of .001 matched to the nearest neighbour, including controls with tied propensity scores, and restrict the treated to the region of common support. Varying these propensity score matching parameters did not substantially change our results. Propensity score matching was done in Stata (version 13.1) using psmatch2.

Table 5 shows the point estimates and 95 per cent CIs for the three models for all five outcomes at the national level. None of the matching algorithms changes the interpretation of the authors' results, although the effect sizes for the reproductive health coverage indicators are slightly decreased in models 2 and 3. The effect on perinatal deaths and neonatal deaths are increased in models 2 and 3.

Table 5: National-level results of propensity score matching, estimated treatment effect[a] and 95 per cent CI

	Lim *et al.* (2010) Table 2 results (exact matching)	Model 1[b]	Model 2[c]	Model 3[d]
Antenatal care, three visits	10.7% (9.1 to 12.3)	10.9% (10.1 to 11.8)	9.4% (8.2 to 10.6)	9.2% (7.9 to 10.5)
In-facility births	43.5% (42.5 to 44.6)	46.8% (46.1 to 47.5)	40.8% (39.6 to 42.0)	42.9% (41.7 to 44.1)
Skilled birth attendance	36.6% (35.6 to 37.7)	40.8% (40.1 to 41.5)	33.7% (32.5 to 35)	35.3% (34.1 to 36.5)
Perinatal deaths (per 1,000 pregnancies)	-3.7 (-5.2 to -2.2)	-3.8 (-5.1 to -2.5)	-4.0 (-6.3 to -1.7)	-3.9 (-6.7 to -1.2)
Neonatal deaths (per 1,000 live births)	-2.3 (-3.7 to -0.9)	-2.3 (-3.6 to -1)	-4.3 (-7 to -1.5)	-5.0 (-8 to -1.9)

Note: [a] Treatment effects are expressed as the change in predicted probabilities.
[b] Covariates include BPL card, urban, state and coarsened wealth group, caste group, mother's education group, number births groups and mother's age group.
[c] Covariates include district survey weights, residential category, district mean income, birth interval, multiple birth, religion, state and uncoarsened wealth group, caste group, mother's education group, number births groups and mother's age group.
[d] Model 2 + district fixed effects in matching algorithm and logistic regression.

We repeated the above analyses for high-focus states separately (Table 6). We find, similar to the authors' results, that effect sizes for in-facility birth and skilled birth attendance outcomes remained higher in high-focus states compared with the national-level results. However, we find insignificant effects for mortality outcomes in the high-focus states. Due to smaller sample size and the incorporation of survey weights in the propensity score algorithm, the CIs of our models 2 and 3 are quite large for mortality outcomes.

Generally, we find that the propensity score matching and coarsened exact matching results are quite similar. This is not surprising: both methods are methodologically doing similar things – trying to match individuals that are similar on a number of observable characteristics, with the goal of reducing selection bias. In this case, the coarsening of variables in coarsened exact matching does not have an important influence on which individuals are matched, and neither does adding more covariates in the propensity score matching. This may be because the region of common support (the overlap of probabilities of being treated, for both treated and untreated individuals) is large (see Appendix Figure 1). Thus, it is not difficult for either algorithm to find appropriate matches, leading to very similar results.

Table 6: Results of propensity score matching for high-focus states, estimated treatment effect[a] and 95 per cent CI

	Lim *et al.* (2010) Table 3 (exact matching)	Model 1[b]	Model 2[c]	Model 3[d]
Antenatal care, three visits	11.1% (9.9 to 12.2)	9.8% (8.9 to 10.8)	10.1% (8.5 to 11.7)	9.1% (7.5 to 10.7)
In-facility births	63.8% (63.0 to 64.6)	66% (65.1 to 66.8)	61.9% (60.6 to 63.3)	63.7% (62.3 to 65.1)
Skilled birth attendance	58.7% (58.0 to 59.5)	62.2% (61.4 to 63.0)	54.6% (53.3 to 56.0)	57.0% (55.5 to 58.4)
Perinatal deaths (per 1,000 pregnancies)	-2.9 (-5.1 to -0.6)	-3.3 (-5.7 to -0.1)	-2.5 (-6.4 to 1.4)	-3.5 (-7.7 to 0.8)
Neonatal deaths (per 1,000 live births)	-1.4 (-3.5 to -0.7)	-1.3 (-3.7 to 1.2)	1.3 (-2.9 to 5.5)	-0.8 (-4.8 to 3.2)

Note: [a] Treatment effects are expressed as the change in predicted probabilities.
[b] Covariates include BPL card, urban, state, and coarsened wealth group, caste group, mother's education group, number births groups and mother's age group.
[c] Covariates include district survey weights, residential category, district mean income, birth interval, multiple birth, religion, state and uncoarsened wealth group, caste group, mother's education group, number births groups and mother's age group.
[d] Model 2 + district fixed effects.

4.2 Additional robustness checks

We conducted sensitivity tests of the results to variations in the regression specification. We added additional regressors, including an indicator for having a BPL card and an indicator for whether the woman was facilitated or motivated to go to a facility by an ASHA representative. We also included interaction effects between maternal age and maternal education, because the effect of maternal age on reproductive health outcomes could vary by education (women with more education begin having children later). Finally, we included interaction effects between residential category and district mean income, because there may be differential effects of income for those in urban and rural areas (income may go further in rural areas). None of the sensitivity checks significantly changed the authors' results for the national level or the high-focus states (see Appendix Tables 3 and 4).

Since matching results can vary by the inclusion or exclusion of matching variables, we also tried excluding several of the authors' matching variables in the coarsened exact matching analysis. Characteristics selected as matching variables should be those that could affect selection in JSY and the outcome. The authors matched on an indicator for having a BPL card, urban status, wealth group, caste group, education group, number of births group and age group. We excluded urban status, wealth and caste, hypothesizing that these variables less directly affect JSY uptake and reproductive health outcomes. We were left with matching on BPL card, education, number of births and age. The results, shown in Appendix Table 3, are robust to the authors' findings. The mortality results are stronger, reflecting the possibility that there is likely some confounding by the variables that we excluded. We note that the matching variables selected by Lim *et al.* are consistent with control variables in other literature evaluating JSY (Mazumdar, Mills and Powell-Jackson 2012; Mohanan *et al.* 2013).

4.3 Accounting for differential programme implementation

4.3.1 Motivation

Uneven implementation of the JSY programme across districts can lead to a problem of endogeneity in the treatment variable (Mazumdar, Mills and Powell-Jackson 2011). For example, it is possible that districts with greater JSY coverage were more likely to be effective in other ways that affect health outcomes. There may be omitted variable bias, in that even after matching for observable characteristics, as Lim *at al.* (2010) did, there remain significant unobservable district-level characteristics, such as management ability and capacity of district health authorities.

Lim *et al.* (2010) calculated JSY uptake and national-level means using survey data restricted to births in the 12 months before the interview, in order to avoid periods of differential implementation of the scheme and biases in the comparison of low-fertility and high-fertility areas. Although it does not ensure the scheme had been implemented to the same extent across all states and districts, restricting the data to only the most recent births allows more time for the scheme to have been implemented since it was introduced in 2005. It also reduces, although it does not eliminate, potential biases from differences across low-fertility and high-fertility areas. However, the treatment effects of JSY on the outcomes of interest were estimated on the full, unrestricted dataset.

The analysis by Mazumdar, Mills and Powell-Jackson (2011) employs a differences-in-differences estimation that exploits the heterogeneity in the timing of JSY introduction across districts. They identify the year in which JSY was first introduced in a given district and use this as an indicator for coverage of JSY. Interactions between year of birth and district-level characteristics (such as share of the population below the poverty line and tribal population share) were included to control for potential sources of endogeneity in the timing of JSY introduction. Mazumdar, Mills and Powell-Jackson's instrumental variable strategy leads to an impact parameter that can be interpreted as the effect of JSY at full coverage. The strength of their approach is that it controls for time-invariant unobservables at the district level that may influence study outcomes and be correlated with the introduction of JSY. Unlike the Lim *et al.* (2010) district-level differences-in-differences approach, Mazumdar, Mills and Powell-Jackson's use of individual-level data allows for greater power to estimate the effect size for the health outcomes (excluding maternal mortality).

4.3.2 Methods

We do not repeat what Mazumdar, Mills and Powell-Jackson have already done, instead addressing the differential implementation of JSY across districts through several different approaches. First, we include district as a matching covariate in the propensity score matching analysis and re-estimated results (see section 4.2). Although this would not have been possible in the exact matching approach due to limitations in bin sizes, it can help control for time-invariant unobservable differences across districts that could be related to the scale-up of JSY. We then explored ways to restrict the data before running the analysis.

First, we re-ran the main analysis restricting the data to births in the 12 months prior to the survey. Next, relying on methods used by Mazumdar, Mills and Powell-Jackson (2011) to explore the heterogeneity in the timing of JSY introduction across districts using facility-level data, we re-estimated the results, restricted to districts that had introduced the programme during the three-year period following the start of JSY (2005–2008). This can help reduce biases related to unobservable variables that may influence study outcomes and be correlated with the introduction of JSY during the study period.

Mazumdar, Mills and Powell-Jackson defined the year JSY was introduced in a given district as the first year in which the proportion of (eligible)[1] women giving birth in a public facility who received a cash payment was 10 percentage points greater than the 2004–2005 level. For each district, we used this definition to identify the first year JSY had been introduced. As they did, we also used fiscal years (1 April to 31 March) rather than calendar years, to align more closely with the government's budgetary cycle. Through this approach, we found that of 601 total districts, 183 introduced JSY in 2005–2006, 214 did so in 2006–2007 and another 90 districts introduced JSY in 2007–2008. Overall, 487 districts introduced the programme during the three-year period following the start of JSY, and 108 districts did not. These numbers differ from the Mazumdar, Mills and Powell-Jackson analysis, which found 424 districts had started the programme during the study period (157 by 2005–2006, 156 more by 2006–2007 and 111 more by 2007–2008), and 163 districts did not.[2] Finally, we conducted an analysis that both restricted the sample to births in the last 12 months prior to the survey interview and only included data from district-years in which JSY had been introduced, according to the definition established by Mazumdar, Mills and Powell-Jackson (2011).

4.3.3 Results

Results from the analysis restricted to births in the 12 months prior to the survey were similar to the unrestricted analysis for the reproductive health coverage outcomes (see Table 7). The treatment effect sizes for the health outcomes, however, were much stronger than the base case analysis. Furthermore, for the perinatal mortality outcome, the estimated treatment effect of JSY is higher for high-focus states than other states (although with overlapping 95 per cent CIs).

[1] It is unclear whether their definition of eligible women refers to all women delivering in facilities or only women who meet JSY eligibility requirements. We explored both possible definitions and found a difference of six districts in the total number that had introduced JSY over the study period. We use the latter definition in this analysis.

[2] There are a few possible reasons for this discrepancy. First, they included 587 districts that were consistently defined across the DLHS-2 and DLHS-3 datasets, whereas we include data for 601 districts as defined in DLHS-3. Second, their analysis also included women 45 to 49 years. Finally, they defined eligibility in non–high-focus states as women from households living below the poverty line, from scheduled castes and tribes ***or*** with two or fewer live births, as opposed to our definition in which eligible women were those from households living below the poverty line, from scheduled castes and tribes, ***and*** those with two or fewer live births. Our definition was based on government documents stating that women from non–high-focus states were only eligible for the cash assistance for their first two live births (MHFW 2006).

Table 7: Estimated treatment effect[a] nationally and by high-focus, north-east and other states restricted to births in the 12 months prior to the survey

	National (N=61,366[b]) Mean (95% CI)	High-focus states (N=42,627[b]) Mean (95% CI)	North-east states (N=2,457[b]) Mean (95% CI)	Other states (N=16,282[b]) Mean (95% CI)
Antenatal care, three visits	10.0% (8.6, 11.3)	8.9% (7.5, 10.4)	16.0% (10.5, 21.5)	3.5% (2.2, 4.8)
In-facility births	48.9% (47.9, 49.8)	66.8% (65.9, 67.7)	34.2% (28.6, 39.9)	8.9% (7.4, 10.3)
Skilled birth attendance	42.6% (41.6, 43.5)	61.6% (60.7, 62.5)	27.6% (23.0, 32.2)	7.1% (5.9, 8.4)
Perinatal deaths (per 1,000 pregnancies)	-12.6 (-15.4, -9.7)	-13.1 (-16.8, -9.4)	-7.5 (-12.6, -2.4)	-11.2 (-15.8, -6.6)
Neonatal deaths (per 1,000 live births)	-9.9 (-12.9, -7.0)	-9.9 (-13.9, -6.0)	-1.2 (-2.8, 0.3)	-11.2 (-16.2, -6.3)

Note: [a] Treatment effects are expressed as the change in predicted probabilities
[b] In-facility births outcome

We find that restricting the data to district-years that had introduced JSY – according to the way Mazumdar, Mills and Powell-Jackson (2011) defined this – resulted in stronger estimated treatment effect sizes for nearly all outcomes (see Table 8). With the exception of antenatal care, the effect sizes for reproductive health coverage indicators are slightly increased. Again, the effect on perinatal deaths and neonatal deaths were stronger than in the unrestricted analysis. The corresponding trends across high-focus, north-east and other states were similar to the Lim *et al.* findings (see Table 8).

Table 8: Estimated treatment effect[a] nationally and by high-focus, north-east and other states restricted to district-years that had introduced JSY, as defined by Mazumdar, Mills and Powell-Jackson (2011)

	National (N=107,555[b]) Mean (95% CI)	High-focus states (N=71,940[b]) Mean (95% CI)	North-east states (N=4,985[b]) Mean (95% CI)	Other states (N=30,630[b]) Mean (95% CI)
Antenatal care, three visits	10.8% (9.8, 11.8)	10.2% (9.1, 11.4)	16.0% (11.4, 20.6)	3.9% (2.9, 4.9)
In-facility births	48.0% (47.1, 48.9)	67.1% (66.3, 67.9)	38.2% (33.9, 42.6)	8.1% (6.9, 9.3)
Skilled birth attendance	41.6% (40.6, 42.5)	62.3% (61.5, 63.0)	33.4% (29.5, 37.4)	6.5% (5.5, 7.5)
Perinatal deaths (per 1,000 pregnancies)	-7.1 (-8.9, -5.3)	-6.3 (-8.9, -3.7)	-4.1 (-7.0, -1.2)	-8.8 (-11.4, -6.2)
Neonatal deaths (per 1,000 live births)	-5.3 (-7.1, -3.5)	-4.1 (-6.8, -1.4)	-1.6 (-4.7, 1.4)	-8.4 (-10.9, -6.0)

Note: [a] Treatment effects are expressed as the change in predicted probabilities
[b] In-facility births outcome

The final analysis, restricting both births in the 12 months prior to the survey and district-years that had implemented JSY dropped roughly half of all births nationally and across all groups of states over which we ran the analysis. In general, we found similar results as the earlier analysis restricted to births in the last 12 months without considering the potential district-level introduction of JSY (Table 9). With the exception of the north-east states, treatment effect sizes for health outcomes were even stronger than the previous model specification, without restricting to births in the last 12 months. These results also indicate that JSY had a stronger effect (although not statistically different) on reducing perinatal mortality in high-focus states than it did in other states.

This is in contrast to previous findings by the authors and our pure replication efforts, showing a smaller effect of the programme on health outcomes in high-focus states compared with other states (non–high-focus and non–north-east states).

Table 9: Estimated treatment effect[a] nationally and by high-focus, north-east and other states, restricted to births in the 12 months prior to the survey and district-years that had introduced JSY, as defined by Mazumdar, Mills and Powell-Jackson (2011)

	National (N=51,820[b]) Mean (95% CI)	High-focus states (N=37,608[b]) Mean (95% CI)	North-east states (N=1,573[b]) Mean (95% CI)	Other states (N=12,639[b]) Mean (95% CI)
Antenatal care, three visits	10.3% (8.9, 11.7)	9.1% (7.7, 10.6)	17.0% (10.3, 23.7)	3.1% (1.8, 4.5)
In-facility births	50.6% (49.5, 51.6)	67.6% (66.7, 68.6)	39.1% (32.8, 45.4)	8.6% (6.9, 10.2)
Skilled birth attendance	44.4% (43.3, 45.5)	62.4% (61.5, 63.3)	32.7% (26.8, 38.7)	7.0% (5.6, 8.4)
Perinatal deaths (per 1,000 pregnancies)	-13.9 (-16.9, -10.8)	-14.3 (-18.2, -10.4)	-7.2 (-13.1, -1.4)	-12.6 (-17.6, -7.7)
Neonatal deaths (per 1,000 live births)	-11.3 (-14.5, -8.2)	-11.8 (-15.9, -7.6)	-1.6 (-3.7, 0.4)	-11.8 (-17.2, -6.5)

Note: [a] Treatment effects are expressed as the change in predicted probabilities
[b] In-facility births outcome

4.4 State level heterogeneity

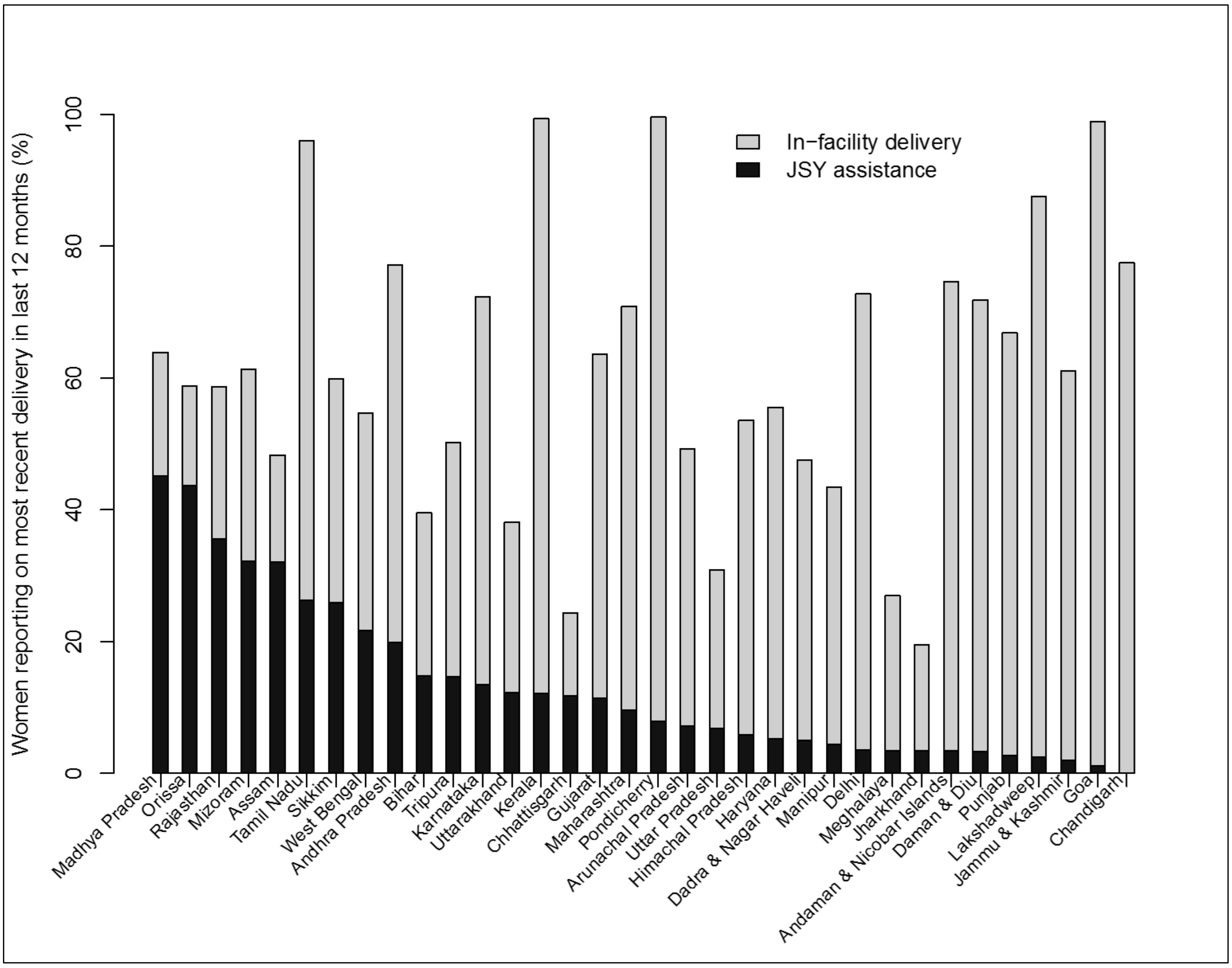
4.4.1 Motivation

Implementation of the JSY programme varied considerably across states. Variation in the programme included differential eligibility guidelines, amounts disbursed to women and payment processes. Additionally, state-to-state variety in their implementation and promotion of the programme may have led to disparities in awareness of JSY's existence. Finally, physical and cultural barriers in remote areas may have contributed to differential uptake of the programme. Lim *et al.* (2010) document issues of differential eligibility, implementation, and uptake across states. They conducted state-specific regressions for states with sufficient sample size and found that the effect of JSY on in-facility delivery and skilled birth attendance varied greatly by state. They also found that the state-specific regressions of JSY on health outcomes could not be assessed because of the small sample sizes. However, they did not show the state-specific results for coverage outcomes, presenting only pooled treatment effects for the national level, high-focus states, north-east states and other states in their final analysis.

Given the size and heterogeneity of many Indian states, variation in treatment effects across geographical areas can be quite important. Understanding which states succeed most in improving outcomes, and why, can guide policymakers and local governments towards better decisions about where and how to invest their resources. For example, a recent article on the World Bank blog discussed poor conditions and overworked staff in community health centres (Das and Hammer 2014). Without basic infrastructure and adequate quality of care necessary to accommodate more institutional deliveries, mortality outcomes will not improve, and may worsen, even as coverage of facility-based deliveries increases. In the following analysis, we attempt to identify which states were most effective at achieving the goals of the JSY programme.

We begin by demonstrating the importance of heterogeneity between states by again replicating the authors' Figure 2, which shows the percentage of women reporting receipt of financial assistance from JSY. This time we add the total percentage of in-facility delivery for each state (see Figure 1). We can see that although states towards the right hand side of the x-axis reported low JSY assistance, a large percentage of women who gave birth in those states were delivering in health facilities. For example, of the 10 states with the lowest percentage of women reporting receipt of financial assistance from JSY, 8 had rates of in-facility delivery higher than 60 per cent. In many of these states, we may not expect the JSY programme to have a significant effect on in-facility delivery. Presenting national or group-level results alone may obscure important variation at the state level by averaging results across states.

Figure 1: Percentage of women reporting receipt of financial assistance from JSY and percentage of women reporting an in-facility delivery, among all women who gave birth in past 12 months, by state



4.4.2 Methods

We investigate the state-specific effects to better illustrate the substantial variation among states and to identify the characteristics of states that had positive treatment effects than those with null or negative treatment effects. We conduct the analysis using the Final Working Dataset and match on the coarsened survey weights using coarsened exact matching, since the propensity score matching and coarsened exact matching yielded similar results. We estimate multilevel linear probability models since the marginal effects of the logistic regressions are similar, with varying intercepts and varying slopes for the JSY coefficient. For all analyses, we incorporate state survey weights to make the samples representative at the state level.

We first conduct a completely unpooled analysis using state-specific regressions for each outcome. The results are in Appendix Table 5. The state sizes range from 39 observations to 29,000 observations. The median state size is 1,910 observations. Because of this, many of the small states are very imprecisely estimated, with standard errors larger than the point estimates. Some state effects could not be estimated due to lack of variation in the outcome at the state level.

To improve the state-specific estimates, we estimate a multilevel random effects model. The multilevel random effects model is effectively a compromise between the completely unpooled and the fully pooled models. Formally, a model with random effects estimates the unit-specific mean as a weighted average of the pooled estimate and the unit-specific estimate of the mean. Our units are the Indian states, where we believe exists important heterogeneity. The weights that contribute to the weighted average are the precisions of the pooled estimate and the state-specific estimate. Because some of the states have small sample sizes, a completely unpooled fixed effects analysis will yield highly variable estimates for those states. With the random effects model, we borrow strength across states to improve individual state estimates. For larger states, we have good estimates, while for smaller states, we borrow information from other states to obtain more accurate estimates. In other words, we allow the treatment effect for the state to incorporate information from the overall mean, calculated from the data of the other states. The states with small sample sizes are 'shrunk' towards the overall pooled mean. This framework allows us to deal with the cross-state heterogeneity (Gelman *et al.* 2004, Gelman and Hill 2006).

The varying intercept, varying slope model is as follows,

$$(1)\ y_{ij} \sim N\left(\alpha_{j[i]} + \beta_{j[i]} JSY_i + \gamma X_i, \sigma_y^2\right), for\ i = 1, \dots, n$$

$$(2)\ \begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix}, \begin{pmatrix} \sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{pmatrix}\right), for\ j = 1, \dots, J$$

where $\alpha_{j[i]}$ is the intercept for individual *i* in state *j* and $\beta_{j[i]}$ is the slope on JSY for individual *i* in state *j* (equation 1). X_i is a matrix of covariates, identical to the one the authors use. The state intercepts and slopes are drawn from a multivariate normal distribution that includes a between-group correlation parameter ρ (equation 2).

The rationale for using multilevel modelling is that it allows us to study effects that vary by group. In classical regression using only local information, estimates of varying effects can be noisy when there are few observations per group (or impossible to estimate, as we find in our unpooled analysis). In the unpooled analysis, the data is overfitted in each state due to the small sample size. It overstates the variation among states and tends to make individual states look more different than they actually are (Hunter and Schmidt 2000). Intuitively, it would be strange if states had uniquely determined treatment effects, whereby information from one state would be utterly uninformative for another state; borrowing information across states makes sense because, despite variation in implementation and eligibility, JSY operates in a similar way across states. Multilevel modelling allows the estimation of group averages and group-level effects as a compromise between the oversimplified pooled analysis and the noisy within-group estimates of the unpooled analyses (Gelman and Hill 2006).

There are two additional assumptions associated with multilevel modelling, as compared to a pooled regression that includes state-level indicators (fixed effects). First, we assume the random effects are normally distributed. Second, we assume the individual specific effects are uncorrelated with the independent variables. In the first case, that assumption is also made with fixed effects, except that the variance across the fixed effects is set to infinity. That is, the random effects specification models the varying intercepts and slopes as arising from a distribution with finite and estimable variances σ_{α}^2 and σ_{β}^2, whereas the fixed-effects specification assumes the intercepts are distributed with infinite variance (Clark and Linzer 2015; Gelman and Hill 2006). Regardless of whether the random effects actually match the data generating process, the random-effects model can still be specified and may be preferable to a fixed-effects specification, due to the large variance mentioned above (Clark and Linzer 2015; Greene 2008).

The second assumption is more disputable, because we may introduce bias if the assumption is violated. In multilevel models, bias is introduced if there is correlation between the random effects and the independent covariates. This is in essence an omitted variables problem; there is some variable that predicts the outcome but is not included in the model, and, as a result, the higher or lower levels of *y* in unit *j* due to this variable are accounted for instead by the random effect (Clark and Linzer 2015). Other research has found via simulations that even in the presence of violations of that assumption, the random-effects estimator can still be preferable or no worse than the fixed-effects estimator (Clark and Linzer 2015). With larger amounts of data, there is no discernible difference in estimates of β between the two estimators, even when the regressor and random effects are correlated (Clark and Linzer 2015).

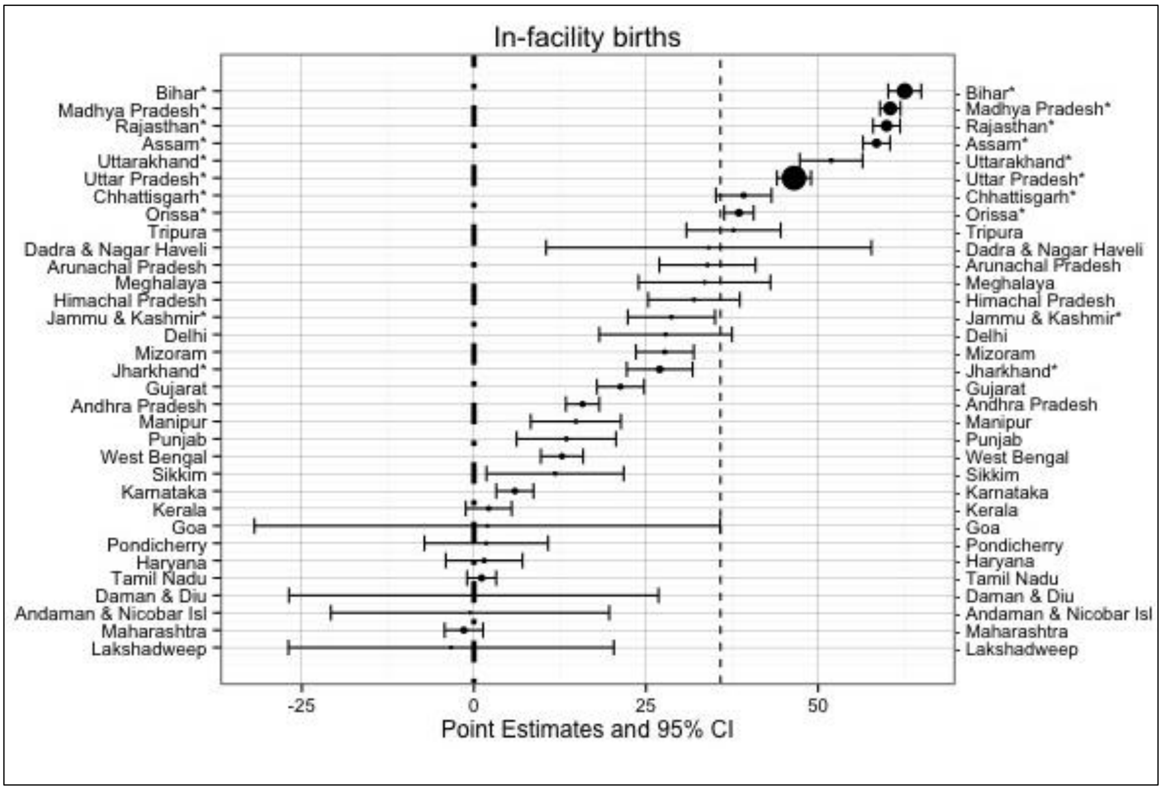
4.4.3 Results

We find that the random-effects model yields slightly more precise estimates, but they are still quite poor for the smaller states. However, we find interesting variation across states. Figure 2 shows the state random effects and 95 per cent CI for in-facility births, compared with the national pooled average of 35.8 percentage points (dotted line).

Asterisks by the names of the states indicate high-focus states. The thicker vertical dot-dash line is the null, implying no effect of JSY on deliveries in a health facility. The size of the point estimate corresponds to the sample size of the state.

There is substantial variation across states, with larger states generally having greater effects of JSY on in-facility births. Most of these larger states are also high-focus states. This is consistent with the authors' findings that high-focus states had larger effects of JSY on in-facility births than non–high-focus states. This is also consistent with the idea that some states already had very high levels of institutional delivery, near 100 per cent (such as in Tamil Nadu or Pondicherry), and so had little room for improvement. Thus, receipt of JSY was associated with no change in rates of institutional delivery for those states, whereas in Bihar, a state with low facility-based delivery, receipt of JSY increased the probability of delivering in a facility by 55 percentage points. Most states have point estimates significantly different from the null. The trend shown in Figure 2 is similar for skilled birth attendance (pooled national average of 50.8 percentage points) (Appendix Figure 5). There is much less variation across states in the effect of JSY on antenatal care visits (pooled national average of 8.6 percentage points), and high-focus states do not have a larger effect size than the other states (see Appendix Figure 4).

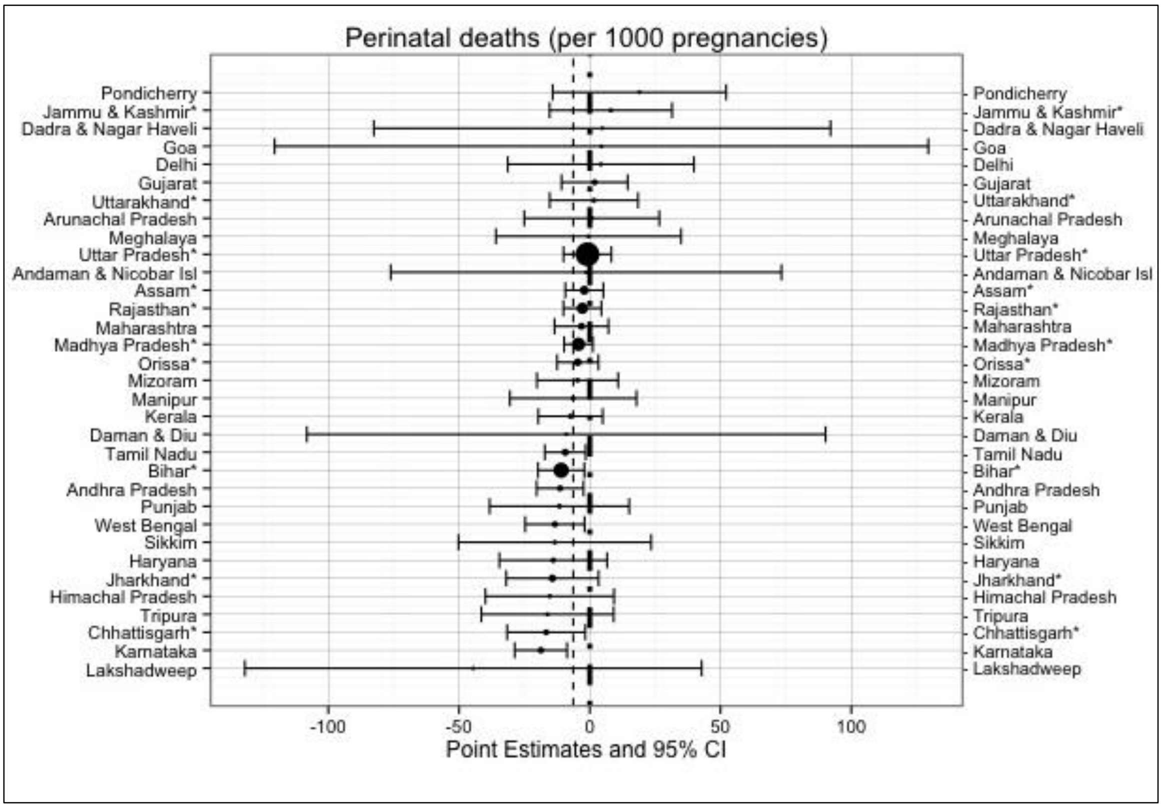
Figure 2: Point estimates and 95 per cent CIs for state random effects on in-facility birth



Note: Size of point estimate relative to size of state sample. Vertical dashed line is national-level pooled average. Thick vertical dot-dash line is 0. High-focus states indicated with asterisks.

The results on health outcomes look a bit different. Figure 3 shows the same figure for perinatal deaths, with a pooled national average of –6.3 per 1,000 pregnancies and much less variation in the point estimates. Due to small sample size and use of survey weights in the matching analysis, these estimates may suffer from a lack of power in detecting significant effects. The resulting confidence are quite large for some states that had low uptake of JSY and low occurrence of reported neonatal deaths. Only the states of West Bengal, Karnataka, Bihar, Tamil Nadu, Chhattisgarh, and Andhra Pradesh have effects significantly different from 0; of these states, only Bihar and Chhattisgarh are high-focus states. On the other hand, some large states like Uttar Pradesh were conceivably large enough to see an effect, and the estimate of the effect of JSY was essentially null, demonstrating that despite a large increase in institutional delivery, JSY did not improve perinatal health outcomes.

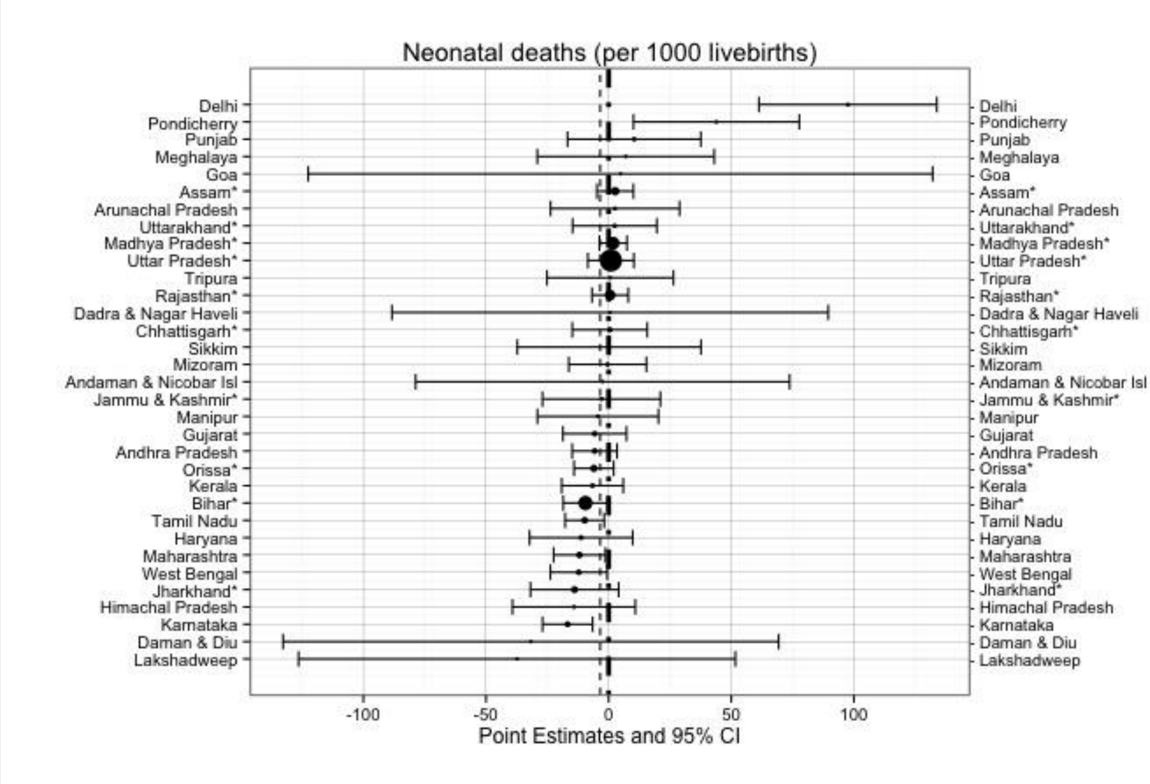
Figure 3: Point estimates and 95 per cent CIs for state random effects on perinatal deaths



Note: Size of point estimate relative to size of state sample. Vertical dashed line is pooled average. Thick vertical dot-dash line is 0. High-focus states are indicated with asterisks.

Results were similar for neonatal deaths (pooled national average –3.5 per 1,000 live births) (Figure 4). States with significantly negative effects for neonatal deaths include Bihar, Karnataka, Maharashtra, Tamil Nadu and West Bengal. Only Bihar is a high-focus state. In two states, Delhi and Pondicherry, we found positive effects of JSY on neonatal deaths. It is unclear why we find positive effects in these states. One possibility is that there may have been especially poor data quality in those states. Lim *et al.* found that estimates of the number of perinatal and neonatal mortalities from the DLHS were about 13 per cent lower than those from the national family health surveys. The CIs in Delhi and Pondicherry are very large, so we know the effects are imprecisely estimated. Alternatively, a lack of quality or understaffing in health facilities could foreseeably lead to worse health outcomes for women who choose to deliver in a facility. We investigate this further in the next section.

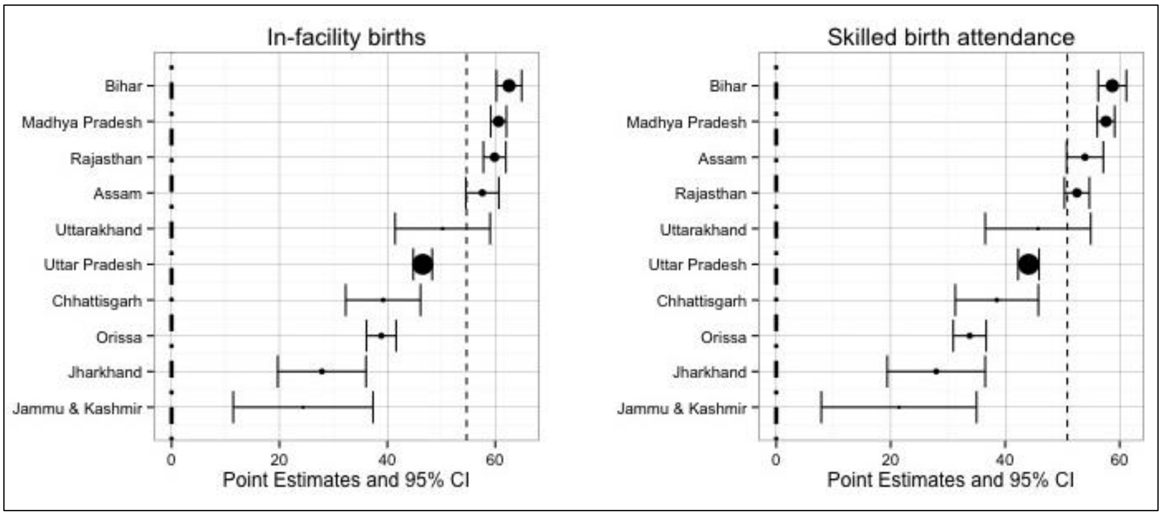
Figure 4: Point estimates and 95 per cent CIs for state random effects relative to overall JSY mean on neonatal deaths



Note: Size of point estimate relative to size of state. Vertical dashed line is pooled average. Thick vertical dot-dash line is 0. High-focus states are indicated with asterisks.

Finally, as found by Lim *et al.*, our results show that the effect of JSY was generally higher for reproductive health coverage indicators in high-focus states than non–high-focus states. Figure 5 shows the effect of JSY on in-facility birth and skilled birth attendance outcomes for high-focus states, with high focus pooled averages of 54.5 and 50.8 percentage points for in-facility birth and skilled birth attendance, respectively. In this analysis, we matched on the state survey weight and reran our random effects model on only the high-focus states. For these outcomes, all high-focus states have significantly positive effects of JSY with small CIs; however, there is wide variation in the magnitude of their effects.

Figure 5: Point estimates and 95 per cent CIs for high-focus state random effects relative to overall JSY mean on in-facility births and skilled birth attendance



Note: Size of point estimate relative to size of state sample. Vertical dashed line is pooled high-focus average. Thick vertical dot-dash line is 0.

In general, average district wealth was negatively associated with significant reproductive health coverage indicators (results not shown). This is consistent with the findings that high-focus states, which are the poorer states, saw the most pronounced effects of JSY on reproductive health coverage indicators.

Although Lim *et al.* (2010) state that the variation in state-specific regressions was mostly accounted for by differences between high-focus and non–high-focus states, we still find substantial variation in the effect of JSY on reproductive health coverage indicators in high-focus states, especially for in-facility birth and skilled birth attendance. There is less variation in the health outcomes, most likely because the health outcome point estimates are quite small overall, so there is less room for variation, along with larger CIs. The varied impact on coverage may also be due to states' varied capacity to implement JSY, timing of implementation and promotion of the programme. Finally, some states were already performing quite well on coverage indicators (as shown in Figure 1), so there was much less room for improvement, whereas other states had a great deal of room to improve.

These results show that any evaluation of the JSY programme should not be limited to the national level. States had considerably varying results from the programme, perhaps due to other state programmes being run at the same time. For example, in 2006, the Gujarat state government rolled out the *Chiranjeevi Yojana* programme, a public-private partnership designed to increase facility-based delivery rates. This programme overlapped with JSY, potentially causing confusion or programme interactions (Mohanan *et al.* 2013). It is important to evaluate the state-level effects with as much precision as possible in order to understand the programme's actual impact.

4.5 Additional descriptive analyses

4.5.1 State capacity

One possible reason for the large variation in state effects may be due to state capacity. We investigate the role of state capacity on treatment effect by comparing our multilevel results for coverage and health outcomes with various state indicators: the Human Development Index, the size of population per hospital bed and the per-capita state expenditures on health. This analysis was added following feedback from a referee for this report.

We use the Human Development Index for each state in India in 2011 (Gandhi *et al.* 2011). The Human Development Index is a composite of three dimensions: life expectancy, mean years of schooling and literacy rate, and monthly per capita expenditure adjusted for inflation and inequality.[3] The correlation between Human Development Index and average in-facility birth is quite strong (.79, p<.001). High-focus states have the highest treatment effects but the lowest fraction of in-facility birth. The results are very similar for the other health coverage and health outcome indicators, and similar when we limit the analysis to high-focus states. However, when we subset to high-focus states (which would have the most room for improvements in in-facility birth) and correlate the JSY treatment effect on in-facility birth with the Human Development Index, we do not find a significant correlation, with similar findings for the other coverage and health outcomes indicators.

We further investigate whether variation in health infrastructure and health spending can explain the variation we see in JSY treatment effects across high-focus states. We correlate population per hospital bed in each state with the health coverage and health outcomes (Choudhury and Nath 2012). We find similar results as above: the population per bed is highly correlated with average health indicators (>.61 for all indicators and significant at the .001 level). However, when we measure the correlation between JSY treatment effects and population per bed for high-focus states, we do not find a significant correlation. Finally, we do the same analysis for per-capita state expenditures on health in 2006 (Kumar *et al.* 2009) and find similar results.

From this analysis, we find that state capacity (as measured by the Human Development Index, population per bed and per capita health expenditures) is correlated with levels of health coverage indicators and health outcomes, but we do not observe a relationship between state capacity and the magnitude of the JSY treatment effect. This analysis is limited by the fact that there are only 10 high-focus states, and by the simple descriptive nature of the analysis. We do find that states with low capacity were appropriately targeted as high-focus, but more research is needed to understand why the effect of JSY on intervention coverage outcomes was greater in some high-focus states than others.

---

[3] Although it is not a perfect measure of state capacity, the Human Development Index is likely to be highly correlated with capacity, because states with better capacity have improved education and work opportunities for their citizens.

4.5.2 Exploring causal pathways

One of the potential limitations of any analysis seeking to evaluate the impact of JSY on institutional delivery is the issue of reverse causality: women receive the cash incentive upon delivering in a health facility (Mazumdar *et al.* 2011). As a result, it is possible that some women who would have had an institutional delivery in the absence of the programme received a cash payment just because they gave birth at a participating facility. Although we are unable to address the issue of reverse causality in this replication exercise, we sought to explore one possible causal pathway to increased antenatal care visits and in-facility delivery as a result of JSY. To this end, we carried out a descriptive analysis on the role of the village-level health activists, ASHAs, in facilitating or motivating women to seek antenatal care and deliver in a health facility.

ASHAs are a key component of the NRHM, who are selected from villages and trained to act as the interface between the community and the public health system (MHFW n.d.). Although states have the flexibility to modify payments to ASHAs for services delivered, ASHAs in most states receive payments for providing referrals and assistance to pregnant women for antenatal care visits and institutional delivery (IIPS 2010). For this analysis, we made use of two survey questions in particular: 'Who facilitated or motivated you to avail of antenatal care?' and 'Who facilitated or motivated you to go to a health facility for delivery?' Both questions were asked before the question about whether women had received financial assistance from JSY for their most recent delivery, making it less likely that responses to the ASHA's question would be influenced by the response to the question about funds from JSY. Because part of the ASHAs' role involves facilitating transport to a facility to give birth, we also used the survey question, 'Who arranged the transportation to take you to the health facility for delivery?' asked of women who delivered in a health facility.

We ran multivariable logistic regressions controlling for individual and household-level characteristics to determine whether women who reported receiving assistance or motivation from an ASHA to seek antenatal care or deliver in a health facility were more likely to report receiving financial assistance from JSY. Our results show that women who reported being motivated to seek antenatal care or deliver in a health facility by ASHAs were most likely to receive payments through the programme, compared with women who reported any other source of assistance or motivation (see Table 10). Women who reported being assisted or motivated to seek antenatal care or deliver in a health facility by an ASHA were more than 2.5 times more likely to have received financial assistance from JSY, compared with women who did not receive motivation or assistance from a health worker. Results were similar for antenatal care and in-facility delivery.

Table 10: Multivariable logistic regression of association between women's receipt of financial assistance from JSY for their most recent birth and reporting assistance or motivation for antenatal care or institutional delivery[a]

	Facilitated or motivated to seek antenatal care (n=166,946)				Facilitated or motivated for in-facility birth (n=122,439)			
	OR	(95% CI)		p-value	OR	(95% CI)		p-value
ASHA	2.80	2.59	3.03	<0.0001	2.64	2.44	2.86	<0.0001
Doctor	1.15	1.09	1.21	<0.0001	1.04	0.98	1.11	0.15
Auxiliary nurse midwife	1.32	1.26	1.39	<0.0001	1.20	1.14	1.27	<0.0001
Other health worker	1.21	1.14	1.29	<0.0001	1.16	1.08	1.24	<0.0001
Other[b]	1.00				1.00			

Note: [a] Covariates include residential category, district mean income, religion, state and wealth decile, caste group, mother's education group, number births groups and mother's age group
[b] Includes NGO, family, friends, self or others

Results were even stronger when we looked at assistance with transport for institutional delivery. Among women who delivered in a health facility, those who reported receiving assistance from ASHAs for transport had the highest odds of receiving JSY payments, compared with those who received assistance from any other source (Table 11). Compared with women who did not receive support for transport from a health worker, women who received assistance from an ASHA were more than four times as likely to have received cash payments through JSY.

Table 11: Multivariable logistic regression of association between women's receipt of financial assistance from JSY for their most recent birth and reporting assistance or motivation for transport to a health facility for institutional delivery[a]

	Arranged transportation for in-facility birth (n=85,458)			
	OR	(95% CI)		p-value
ASHA	4.38	3.72	5.14	<0.0001
Doctor	1.13	0.97	1.31	0.1200
ANM	1.43	1.21	1.68	<0.0001
Other health worker	1.78	1.47	2.16	<0.0001
Other[b]	1.00			

Note: [a] Covariates include residential category, district mean income, religion, state and wealth decile, caste group, mother's education group, number births groups and mother's age group
[b] Includes NGO, family, friends, self or others

Results across all three analyses were similar, but not as strong, when restricting the data to the 12 months before the survey (results not shown). We reran the above analyses restricting the sample to rural areas, because ASHAs are selected to work at the village level (i.e., in rural areas). The results remained consistent with what we had found without restricting the data (results not shown).

5. Discussion

This replication study re-examined recent work by Lim and colleagues, 'India's *Janani Suraksha Yojana*, a conditional cash transfer programme to increase births in health facilities: an impact evaluation'. In the original work, the authors found that receipt of JSY payments was generally higher in the middle bands of wealth in high-focus states and in those with middle levels of education (not the poorest and least educated, which was the target population). They also found that JSY had a significant effect on increasing antenatal care and in-facility births and was associated with a reduction in perinatal and neonatal mortality.

The pure replication aimed to validate the authors' original findings by re-estimating the key coverage and health outcome results using the same assumptions and empirical models the authors had employed. We were able to replicate the results of this study quite closely. Resolving several coding issues we identified did not substantively change the results; however, our final sample sizes were 1.1 to 1.2 times larger than the authors'.

We began the measurement and estimation analysis with alternate specifications of the matching model and examined the robustness of the results under the more commonly used propensity score matching algorithm. We found that the authors' results are robust to changes in the addition and definitions of various covariates. We also found that the authors' national results are very similar to results we obtained using three different propensity score models. The effect on mortality outcomes is increased using a propensity score model with district fixed effects for the national results. The effect is decreased and insignificant for the high-focus states. Effects on reproductive health coverage indicators remain largely unchanged.

Following this, we examined alternative methods to account for district-level differences in degree of implementation by carrying out analyses that explicitly accounted for the implementation of JSY across districts. We explored various ways of restricting the data before running the analyses. We restricted the data to births in the 12 months before the survey and to district-years that had introduced JSY. In both analyses, results for reproductive health coverage outcomes were similar to the unrestricted analysis. However, the treatment effect sizes for health outcomes were stronger than in the unrestricted analysis. Most interestingly, we find that JSY's effect on reducing perinatal and neonatal mortality was stronger in high-focus states than other states. These important findings illustrate the potential dilution of treatment effect on health outcomes in the base case analysis, when including districts that had potentially not yet introduced JSY, or district-years in which the programme was not yet well established. The implication of these findings is that it is possible that JSY's effects may change over time as the programme is rolled out fully across states.

We also investigate the degree of state-level heterogeneity via a multilevel model with varying intercepts and varying slopes for the effect of JSY. In our state-specific analyses, we find substantial variation in the effect of JSY on reproductive health coverage indicators even within high-focus states, especially for in-facility birth and skilled birth attendance. Due to large CIs, there is less variation in the health outcomes. This analysis contributes to the debate by illuminating the vast discrepancies in the effect of JSY across states. It is clear that some states are not performing as well as others (with potentially increased neonatal mortality in Delhi and Pondicherry), which may be the result of poor clinic infrastructure, understaffing and a shift from private clinics to less able public clinics (Paul 2010; Das and Hammer 2014; Mazumdar, Mills and Powell-Jackson 2011). However, we did not find a correlation between state capacity and magnitude of the effect of JSY for high-focus states as an explanation for the variability.

Finally, we investigated one possible pathway through which JSY may be operating, by exploring the role of ASHAs in motivating women to seek antenatal care and deliver in a facility. We find that women who reported being motivated to seek antenatal care or deliver in a health facility by ASHAs were more likely to receive JSY payments than women who reported any other source of assistance or motivation. Similarly, women who reported being assisted by ASHAs for transport to a health facility for delivery were most likely to receive cash payments through JSY. Although these results do not rule out reverse causality, they suggest that ASHAs have contributed to women's receiving three or more antenatal care visits and delivering in a health facility through the JSY programme.

6. Limitations

This replication is subject to many of the same limitations as the original Lim *et al.* (2010) study. First, we are unable to correct for potential biases resulting from selective uptake of the programme that are related to unobservable characteristics. We are also unable to control fully for the programme's differential implementation across districts and states, although we have carried out several analyses to try and minimise the effect on our estimates. Our results are dependent on the quality of the DLHS, which we are unable to evaluate. Of the outcomes we consider in this analysis, the estimates of perinatal and neonatal mortality are likely to be the most sensitive to data quality. Data on perinatal deaths are particularly difficult to collect through household surveys because of difficulties in accurately assessing stillbirths and challenges in obtaining accurate gestational age. Both perinatal and early neonatal mortality can be subject to underreporting, particularly among groups with less contact with the health system or health workers. Lim *et al.* (2010) found estimates of the number of perinatal and neonatal deaths from the DLHS to be approximately 13 per cent lower than those from the national family health surveys, although it is unclear which estimates are more accurate.

Although our results generally support Lim *et al.*'s findings, in constructing a replication study, we suffer from the same methodological weaknesses of the original authors, particularly biases resulting from confounding by unobservable characteristics. The difference-in-differences strategy employed by Mazumdar, Mills and Powell-Jackson (2011) is a stronger analysis than matching, and another study also using a difference-in-differences analysis of a similar programme in one state in India found null effects (Mohanan *et al.* 2013). More research should be conducted to reconcile the different findings. Once data from the new DLHS are publicly available to researchers, it would be useful to include this additional data in an analysis of the effect of JSY. In the future, incorporating an evaluation plan into the development of conditional cash transfer programmes – for example, via randomly staggered rollout – could provide better evidence on such programmes' effectiveness in improving health outcomes.

7. Conclusions

Similar to Lim *et al.* (2010) and Mazumdar, Mills and Powell-Jackson (2011), we conclude that better administration and preparedness of the programme is important before scale-up. Demand-side interventions by the government can be effective in improving access to health services, but simultaneously strengthening quality of care is crucial. Finally, we believe the national picture presented by Lim *et al.* and Mazumdar, Mills and Powell-Jackson masks important state-to-state differences in programme effects.

Ultimately, this replication was intended to verify and examine the robustness of the findings from Lim *et al.* (2010). Conditional cash transfers are poised to make significant contributions to the health of people in developing countries. However, it is important to understand under what conditions such transfers are successful. JSY was one of the one of the world's largest cash incentive programmes for health. Evaluating this programme has vital implications for future interventions and policies to improve health in developing countries. Therefore, ensuring the robustness of the results found by Lim *et al.* (2010) was an important exercise for the future of development policy. We found that the results are replicable and robust to changes in model specifications and analysis. However, we also found meaningful heterogeneity across states and districts. The effect of JSY on health outcomes may be understated when averaged over all states and districts, especially when accounting for the implementation lag across the country.

Appendix

Appendix Table 1: Coding discrepancies identified in Lim *et al.*’s Stata code and modifications made

		Stata Code
Creating early neonatal mortality variable	Authors’ code	***typo in code: b09_==1 should be b08_==1 gen enm = . replace enm = 1 if ((died == 1)&(b08_ == 1)&(b09_ <= 7)) replace enm = 0 if ((died == 1)&(b09_ == 1)&(b09_ > 7))
	Modified code	**Replace last line with the following replace enm = 0 if ((died == 1)&(b08_ == 1)&(b09_ > 7))

Creating date of last birth variable	Authors’ code	***the code misaligns the year and month of last birth - it takes the highest number month, regardless of whether it corresponds to the last birth year.
		
		forvalues i = 1/6 {
		
		replace cmclastbirth = (12*(yearlastbirth - 1900) + v143a_`i') if (((12*(yearlastbirth - 1900) + v143a_`i') >= cmclastbirth)&(v143a_`i' ~= .))
		
		}

	Modified code	**replace with following code forvalues i = 1/6 { *Correspond month of last birth with year of last birth replace cmclastbirthfixed = (12*(yearlastbirth - 1900) + v143a_`i') if ((yearlastbirth == v143b_`i')&(v143a_`i' ~= .)&(cmclastbirthfixed == .))  *and set missing month to Jan replace cmclastbirthfixed = (12*(yearlastbirth - 1900) + 1) if ((yearlastbirth == v143b_`i')&(v143a_`i' == .)&(cmclastbirthfixed == .)) }
Creating number of births variable	Authors’ code	***There are 16 women with v134==0 but have had a live birth post 2004. ***these get set to missing and get deleted from dataset. gen nb = . replace nb = v134 replace nb = . if (nb == 0) label var nb "Parity"
	Modified code	*add the following line replace nb = 1 if v134==0 & v139_1 == 1
Creating birth index variable	Authors’ code	***This sort is not unique; that is, there are multiple observations with the same sort, producing ties. Stata randomly sorts ties gsort + obs - yearbirth – cmcbirth
	Modified code	**This code sorts uniquely, then sorts by setting the seed on the ties in order to make the dataset reproducible gsort + obs - yearbirth - cmcbirth -enm -sb +mult_birth +alive, gen(sort_num) sort sort_num set seed 101 gen random_number=uniform() gsort + obs - yearbirth - cmcbirth + random_number

Appendix Table 2: Replication of Lim *et al.* (2010) Table 1, 'Multivariable logistic regression of association between women's receipt of financial assistance under JSY for their most recent birth and individual characteristics by use of round three of the DLHS (2007–2009)'

	National (n = 182,764)				High-focus states (n = 111,792)				North-east states (n = 12,903)				Other (n = 58,048)			
	OR	**(95% CI)**		**p-value**	**OR**	**(95% CI)**		**p-value**	**OR**	**(95% CI)**		**p-value**	**OR**	**(95% CI)**		**p-value**
Maternal age (years)																
15–19	1.57	1.44	1.72	b	1.88	1.68	2.11	b	1.64	1.12	2.39	0.0102	1.25	1.06	1.47	0.0065
20–24	1.33	1.24	1.42	b	1.43	1.32	1.56	b	1.39	1.09	1.75	0.0067	1.23	1.09	1.40	0.0011
25–29	1.12	1.05	1.19	0.0006	1.15	1.07	1.24	0.0002	1.10	0.89	1.37	0.3582	1.10	0.97	1.24	0.1354
30–34	1.00				1.00				1.00				1.00			
35–39	0.84	0.76	0.94	0.0013	0.85	0.75	0.96	0.0073	0.79	0.58	1.09	0.1576	0.84	0.66	1.06	0.1484
40–44	0.76	0.63	0.92	0.0043	0.76	0.62	0.94	0.0127	0.76	0.42	1.38	0.3694	0.73	0.39	1.35	0.3105
Number of live births																
1 birth	1.00				1.00				1.00				1.00			
2 births	0.86	0.82	0.90	b	0.76	0.71	0.81	b	1.05	0.88	1.26	0.5621	0.94	0.87	1.01	0.1021
3–4 births	0.55	0.52	0.59	b	0.65	0.60	0.69	b	0.50	0.40	0.62	b	0.41	0.37	0.46	b
5 or more	0.54	0.49	0.60	b	0.66	0.59	0.75	b	0.45	0.27	0.76	0.0029	0.23	0.15	0.34	b
Maternal education																
No education	1.00				1.00				1.00				1.00			
1 to 5 years	1.22	1.15	1.29	b	1.25	1.17	1.33	b	1.15	0.87	1.51	0.3256	1.18	1.06	1.31	0.0026
6 to 11 years	1.41	1.34	1.48	b	1.47	1.38	1.56	b	1.45	1.11	1.88	0.0056	1.28	1.16	1.41	b
12 years or more	1.25	1.15	1.36	b	1.46	1.30	1.64	b	1.53	1.09	2.15	0.0148	1.03	0.90	1.19	0.6607
Household wealth																
Poorest decile	1.00				1.00				1.00				1.00			
Decile 2	0.97	0.89	1.05	0.3986	0.94	0.86	1.02	0.1288	1.46	0.41	5.23	0.5630	0.92	0.75	1.13	0.4521
Decile 3	1.05	0.97	1.13	0.2476	1.01	0.93	1.10	0.7378	2.17	0.72	6.57	0.1698	0.95	0.79	1.16	0.6363
Decile 4	1.05	0.97	1.14	0.2471	1.01	0.93	1.11	0.7536	2.14	0.72	6.32	0.1698	0.93	0.77	1.12	0.4506
Decile 5	1.12	1.03	1.21	0.0061	1.06	0.96	1.16	0.2628	3.38	1.15	9.92	0.0268	0.96	0.80	1.16	0.6879

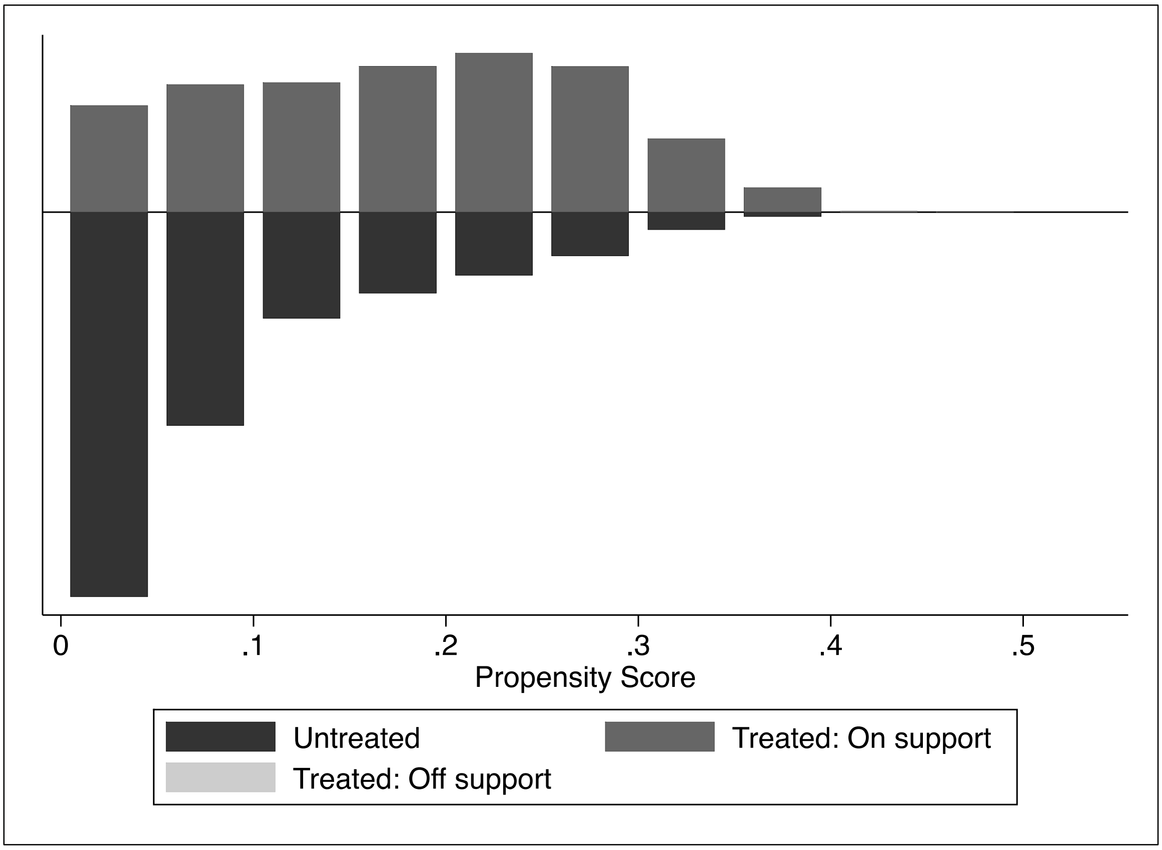
	National (n = 182,764)				High-focus states (n = 111,792)				North-east states (n = 12,903)				Other (n = 58,048)			
	OR	**(95% CI)**		**p-value**	**OR**	**(95% CI)**		**p-value**	**OR**	**(95% CI)**		**p-value**	**OR**	**(95% CI)**		**p-value**
Decile 6	1.12	1.03	1.22	0.0085	1.08	0.98	1.19	0.1339	3.11	1.06	9.14	0.0395	0.93	0.77	1.12	0.4537
Decile 7	1.11	1.02	1.21	0.0167	1.10	0.99	1.22	0.0899	3.69	1.25	10.87	0.0180	0.87	0.72	1.06	0.1638
Decile 8	1.03	0.94	1.13	0.5634	1.00	0.89	1.12	0.9719	3.47	1.18	10.24	0.0242	0.81	0.67	0.99	0.0357
Decile 9	0.72	0.65	0.81	b	0.81	0.71	0.92	0.0018	3.49	1.17	10.42	0.0252	0.49	0.40	0.61	b
Richest decile	0.52	0.46	0.60	b	0.61	0.51	0.72	b	3.79	1.24	11.60	0.0197	0.31	0.24	0.40	b
Caste																
Scheduled caste	1.39	1.31	1.48	b	1.33	1.23	1.44	b	0.92	0.63	1.36	0.6916	1.53	1.38	1.71	b
Scheduled tribe	1.22	1.14	1.30	b	1.13	1.04	1.23	0.0040	0.87	0.62	1.21	0.4013	1.62	1.42	1.85	b
Other backward class	1.14	1.08	1.21	b	1.17	1.10	1.26	b	0.87	0.61	1.22	0.4089	1.20	1.08	1.32	0.0004
Other	1.00				1.00				1.00				1.00			
Religion																
Hindu	1.00				1.00				1.00				1.00			
Muslim	0.75	0.70	0.80	b	0.80	0.74	0.88	b	0.52	0.28	0.98	0.0426	0.75	0.67	0.85	b
Christian	0.89	0.80	1.00	0.0578	0.78	0.62	0.97	0.0290	1.01	0.76	1.34	0.9234	1.00	0.84	1.19	0.9691
Sikh	0.82	0.63	1.06	0.1253	0.43	0.25	0.74	0.0024	1.00	1.00	1.00	.	1.04	0.73	1.48	0.8334
Buddhist	1.12	0.93	1.34	0.2430	3.82	2.21	6.60	b	1.57	1.15	2.14	0.0042	0.79	0.58	1.07	0.1213
Other	0.96	0.77	1.18	0.6885	1.30	0.99	1.71	0.0581	0.69	0.47	1.00	0.0494	1.00	0.53	1.90	0.9947
Urban residence	0.94	0.89	1.00	0.0504	1.11	1.03	1.21	0.0105	1.01	0.82	1.24	0.9263	0.76	0.69	0.83	b
District mean household wealth	0.87	0.82	0.92	b	0.89	0.83	0.96	0.0017	1.31	1.01	1.68	0.0387	0.86	0.78	0.95	0.0026
State																
Andaman & Nicobar Islands	1.56	1.01	2.43	0.0472	(omitted)				(omitted)				0.25	0.16	0.39	b
Andhra Pradesh	6.56	5.85	7.36	b	(omitted)				(omitted)				1.00			
Arunachal Pradesh	1.82	1.47	2.25	b	(omitted)				1.00				(omitted)			
Assam	6.55	5.94	7.22	b	6.87	6.22	7.59	b	(omitted)				(omitted)			

	National (n = 182,764)				High-focus states (n = 111,792)				North-east states (n = 12,903)				Other (n = 58,048)			
	OR	(95% CI)		p-value	OR	(95% CI)		p-value	OR	(95% CI)		p-value	OR	(95% CI)		p-value
Bihar	1.86	1.69	2.06	b	1.91	1.73	2.11	b	(omitted)				(omitted)			
Chandigarh	0.94	0.23	3.93	0.9346	(omitted)				(omitted)				0.22	0.05	0.94	0.0412
Chhattisgarh	1.87	1.63	2.14	b	1.99	1.74	2.28	b	(omitted)				(omitted)			
Dadra & Nagar Haveli	1.77	0.97	3.25	0.0627	(omitted)				(omitted)				0.24	0.13	0.44	b
Daman & Diu	0.69	0.35	1.33	0.2651	(omitted)				(omitted)				0.12	0.06	0.23	b
Delhi	1.39	0.95	2.03	0.0890	(omitted)				(omitted)				0.32	0.22	0.48	b
Goa	1.12	0.50	2.53	0.7786	(omitted)				(omitted)				0.22	0.09	0.49	0.0003
Gujarat	2.51	2.21	2.86	b	(omitted)				(omitted)				0.38	0.33	0.44	b
Haryana	1.40	1.16	1.69	0.0005	(omitted)				(omitted)				0.24	0.20	0.30	b
Himachal Pradesh	1.45	1.15	1.84	0.0019	(omitted)				(omitted)				0.24	0.19	0.31	b
Jammu & Kashmir	0.98	0.77	1.24	0.8680	0.83	0.65	1.07	0.1472	(omitted)				(omitted)			
Jharkhand	0.65	0.55	0.77	b	0.66	0.55	0.78	b	(omitted)				(omitted)			
Karnataka	3.54	3.16	3.95	b	(omitted)				(omitted)				0.56	0.49	0.63	b
Kerala	4.11	3.52	4.82	b	(omitted)				(omitted)				0.70	0.59	0.83	b
Lakshadweep	1.42	0.80	2.52	0.2366	(omitted)				(omitted)				0.21	0.12	0.39	b
Madhya Pradesh	8.89	8.19	9.66	b	9.13	8.41	9.92	b	(omitted)				(omitted)			
Maharashtra	2.14	1.89	2.42	b	(omitted)				(omitted)					0.34	0.30	0.39
Manipur	1.56	1.26	1.93	b	(omitted)				0.89	0.66	1.20	0.4416	(omitted)			
Meghalaya	0.80	0.58	1.11	0.1815	(omitted)				0.67	0.45	1.01	0.0582	(omitted)			
Mizoram	9.29	7.78	11.08	b	(omitted)				4.54	3.57	5.77	<0.0001	(omitted)			
Orissa	7.25	6.61	7.95	b	7.83	7.13	8.60	b	(omitted)				(omitted)			
Pondicherry	4.33	3.35	5.61	b	(omitted)				(omitted)				0.84	0.65	1.10	0.2100
Punjab	0.91	0.69	1.18	0.4694	(omitted)				(omitted)				0.14	0.10	0.19	b
Rajasthan	7.54	6.91	8.23	b	7.75	7.09	8.48	b	(omitted)				(omitted)			
Sikkim	7.38	6.11	8.92	b	(omitted)				3.26	2.48	4.26	b	(omitted)			

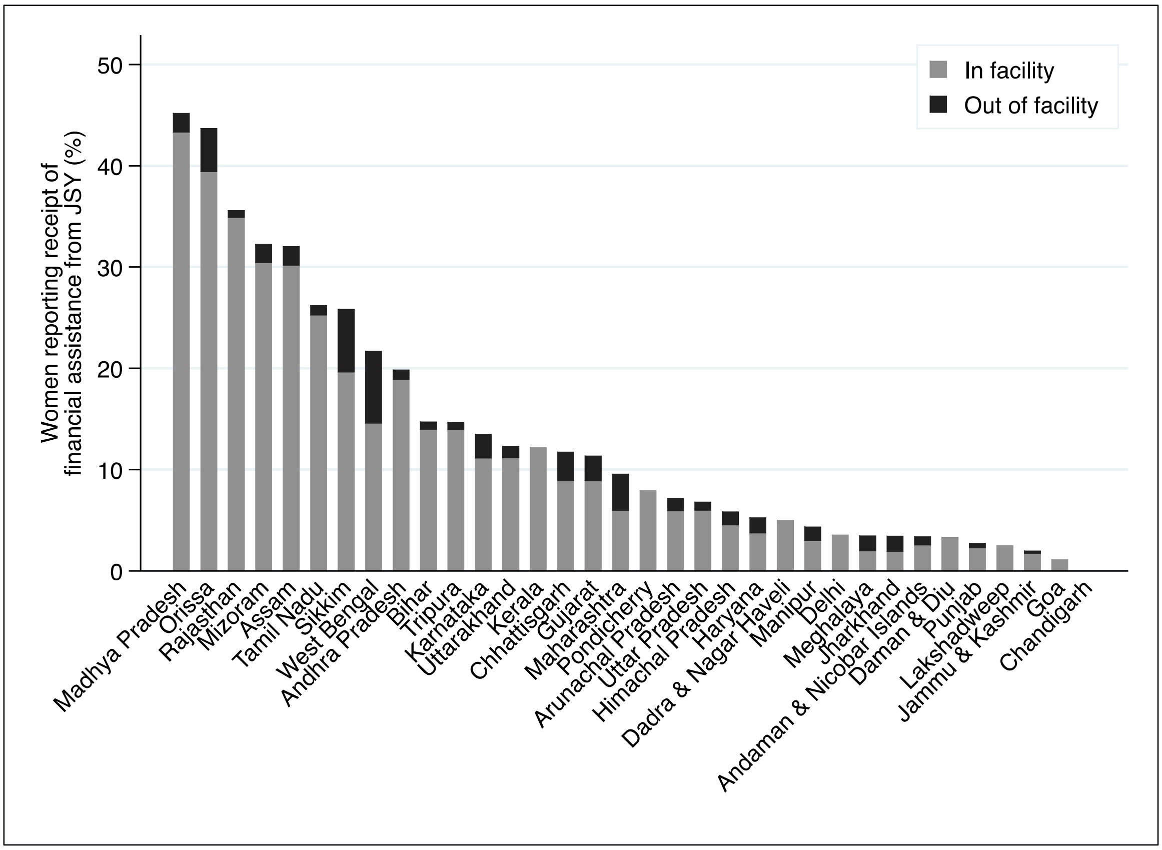
	National (n = 182,764)				High-focus states (n = 111,792)				North-east states (n = 12,903)				Other (n = 58,048)			
	OR	(95% CI)		p-value	OR	(95% CI)		p-value	OR	(95% CI)		p-value	OR	(95% CI)		p-value
Tamil Nadu	7.19	6.44	8.04	b	(omitted)				(omitted)				1.21	1.07	1.36	0.0025
Tripura	2.34	1.87	2.91	b	(omitted)				1.78	1.28	2.49	0.0006	(omitted)			
Uttar Pradesh	1.00				1.00				(omitted)				(omitted)			
Uttarakhand	2.43	2.07	2.86	b	2.46	2.08	2.91	b	(omitted)				(omitted)			
West Bengal	3.37	3.01	3.78	b	(omitted)				(omitted)				0.50	0.43	0.58	b

Note: [a] Using Working Dataset 1, which contains adjusted definition for early neonatal mortality
[b] p-value <0.000

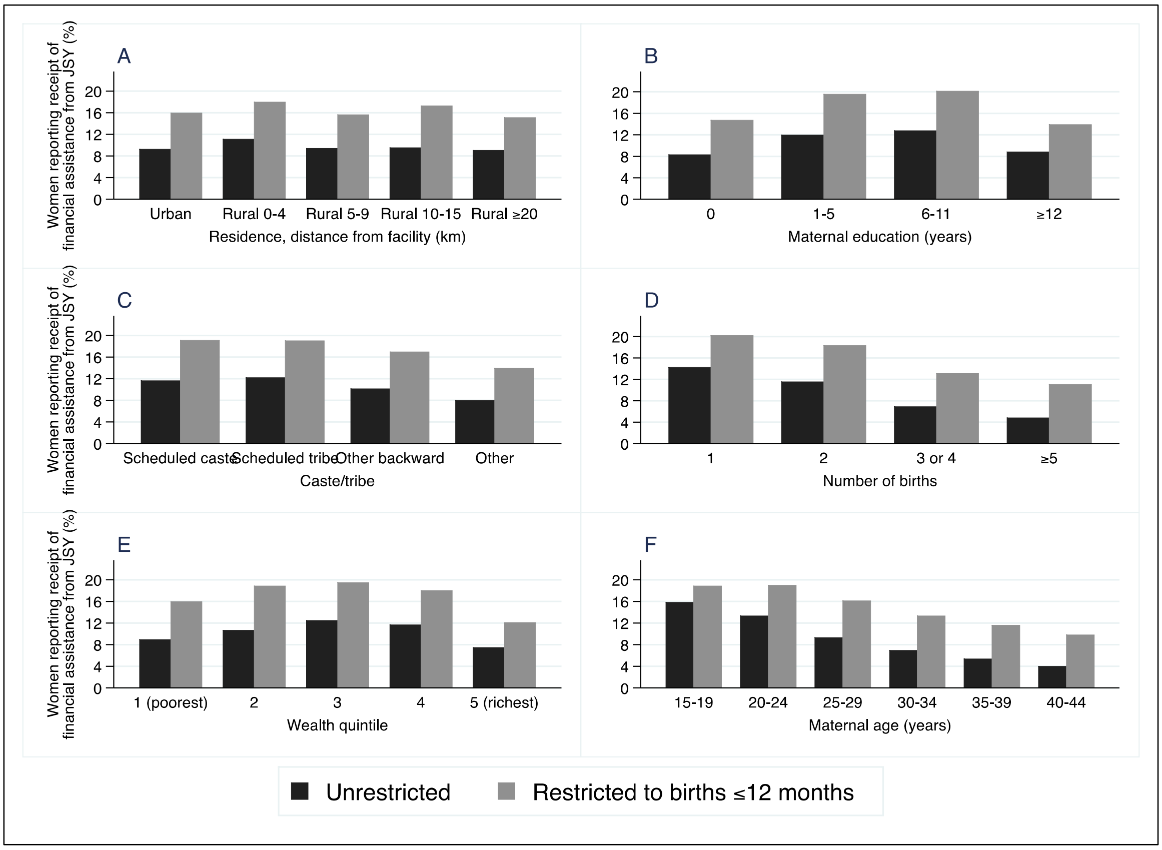
Appendix Figure 1: Region of common support for propensity score matching for results in Table 6, model 1



Appendix Figure 2: Replication of authors' Figure 2, 'Percentage of women reporting receipt of financial assistance from JSY among all women who gave birth in past 12 months by state and location of birth'



Appendix Figure 3: Replication of authors' Figure 3 (using restricted and unrestricted data), 'Percentage of women reporting receipt of financial assistance from *Janani Suraksha Yojana* (JSY) among all women who gave birth in the past 12 months by individual characteristics'



Appendix Table 3: Robustness checks model specifications for national level results

	Additional Covariates[a]	Interaction Effects[b]	Excluding Matching Variables[c]
Antenatal care	10.7 (9.8,11.7)	10.7 (9.8, 11.7)	11.0 (10.1, 11.8)
In-facility birth	43.4 (42.6, 44.2)	43.3 (42.5, 44.1)	43.8 (43.2, 44.5)
Skilled birth attendance	36.7 (35.9, 37.5)	36.5 (35.7, 37.4)	36.9 (36.3, 37.5)
Perinatal deaths	-3.5 (-4.9, -2.2)	-3.57 (-4.9, -2.2)	-3.9 (-5.2, -2.7)
Neonatal deaths	-2.7 (-4.0, -1.4)	-2.7 (-4.0, -1.4)	-2.9 (-4.1, -1.6)

Note: [a] Additional covariates include BPL card and ASHA indicator
[b] Additional covariates plus interaction effects including maternal age x maternal education, and district mean income x residential category
[c] Matching with only BPL card, education, number births, and age (leaving out urban, wealth and caste groups)

Appendix Table 4: Robustness checks for model specifications for high-focus state results

	Additional Covariates[a]	Interaction Effects[b]
Antenatal care	10.92 (9.76,12.09)	10.92 (9.76,12.09)
In-facility birth	63.34 (62.5, 64.17)	63.32 (62.49, 64.15)
Skilled birth attendance	58.49 (57.72, 59.25)	58.39 (57.63, 59.16)
Perinatal deaths	-2.6 (-4.7, -0.4)	Not estimable
Neonatal deaths	-1.6 (-3.7, 0.5)	Not estimable

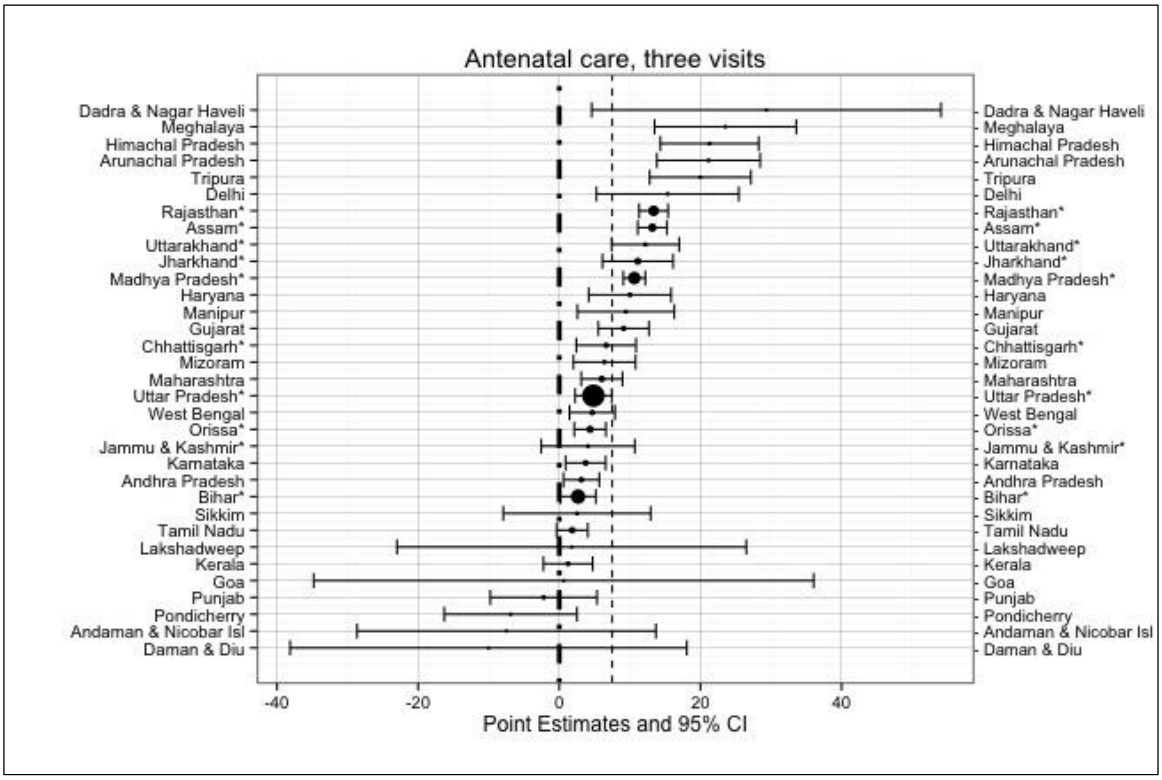
Note: [a] Additional covariates include BPL card and ASHA indicator
[b] Additional covariates plus interaction effects including maternal age x maternal education and district mean income x residential category

Appendix Table 5: State-specific regression estimates for JSY on outcomes

	N	ANC	SE	IFB	SE	SBA	SE	PNM	SE	NNM	SE
Andaman & Nicobar Islands	68	-0.1121	0.0684	-0.0243	0.1393	-0.0243	0.1393	NA	NA	-0.1125	0.0683
Andhra Pradesh	3626	0.0258	0.0092	0.1542	0.0145	0.1393	0.0138	-0.0111	0.0046	0.0261	0.0092
Arunachal Pradesh	1170	0.207	0.041	0.3478	0.039	0.3399	0.0389	-0.0042	0.0044	0.2081	0.041
Assam	7647	0.129	0.0131	0.5837	0.0105	0.547	0.0108	0.0014	0.004	0.1283	0.0131
Bihar	16912	0.0367	0.0119	0.6217	0.011	0.578	0.0116	-0.0109	0.0046	0.0424	0.012
Chhattisgarh	3330	0.0615	0.0259	0.3835	0.018	0.371	0.0219	-0.0168	0.0095	0.0576	0.026
Dadra & Nagar Haveli	39	0.2043	0.2929	0.146	0.2212	-0.0057	0.1663	NA	NA	0.2043	0.2929
Daman & Diu	49	-0.0973	0.1352	0.0949	0.1065	0.0402	0.0663	NA	NA	-0.0973	0.1352
Delhi	872	0.1778	0.0478	0.2686	0.0462	0.2534	0.0455	0.0014	0.0173	0.1783	0.0478
Goa	75	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Gujarat	3271	0.0971	0.0218	0.2049	0.0208	0.1956	0.0204	0.0012	0.0058	0.0977	0.0219
Haryana	2410	0.0909	0.0309	0.0347	0.0314	0.0553	0.03	-0.0154	0.0087	0.0931	0.0309
Himachal Pradesh	922	0.2109	0.0428	0.3241	0.046	0.3213	0.0457	0.0041	0.0054	0.2101	0.0427
Jammu & Kashmir	1175	0.0555	0.0376	0.2565	0.0452	0.2283	0.045	0.009	0.0113	0.0562	0.0381
Jharkhand	5412	0.0958	0.0272	0.264	0.0216	0.2712	0.0248	-0.0099	0.0098	0.0966	0.0272
Karnataka	3970	0.0337	0.0127	0.0479	0.0156	0.0386	0.0146	-0.0205	0.0056	0.0331	0.0128
Kerala	2302	-0.0031	0.0029	0.0117	0.0047	0.0117	0.0047	-0.0045	0.0054	-0.0031	0.0029
Lakshadweep	76	NA	NA	-0.0612	0.0612	-0.0765	0.0506	-0.0234	0.063	NA	NA
Madhya Pradesh	13955	0.1101	0.0087	0.5989	0.0074	0.5718	0.0076	-0.0042	0.003	0.1099	0.0087
Maharashtra	4898	0.0664	0.017	0.003	0.0179	0.0395	0.0176	-0.0043	0.0049	0.0652	0.0171
Manipur	1234	0.1206	0.0333	0.1721	0.0359	0.125	0.0303	-0.0055	0.0051	0.1199	0.0332
Meghalaya	467	0.2443	0.0644	0.3811	0.0599	0.3675	0.0617	0.0081	0.0167	0.244	0.0642
Mizoram	1821	0.0608	0.0187	0.2795	0.0202	0.2311	0.0196	-0.0032	0.0047	0.0603	0.0187
Orissa	5901	0.0335	0.0143	0.3749	0.0124	0.322	0.0126	-0.004	0.0043	0.0378	0.0143
Pondicherry	395	-0.0694	0.0299	0.0007	0.0052	0.0007	0.0052	0.0182	0.0161	-0.0676	0.0303
Punjab	1911	-0.0066	0.0401	0.1442	0.0401	0.071	0.0326	-0.009	0.0132	-0.0071	0.0402
Rajasthan	10938	0.1278	0.0099	0.5927	0.0093	0.5182	0.0096	-0.0018	0.0032	0.1319	0.01
Sikkim	1086	0.0262	0.029	0.1433	0.0308	0.1169	0.0302	-0.0097	0.0076	0.0224	0.0289
Tamil Nadu	4955	0.0088	0.0057	0.0017	0.0067	-0.0044	0.0057	-0.0085	0.0042	0.0083	0.0057
Tripura	617	0.1499	0.047	0.3767	0.0438	0.3574	0.0437	-0.0163	0.0127	0.1443	0.0468
Uttar Pradesh	28975	0.0472	0.0118	0.4683	0.0116	0.4353	0.0123	-0.001	0.005	0.0483	0.0119
Uttarakhand	1638	0.1264	0.0307	0.5104	0.0276	0.4595	0.0286	0.0034	0.0048	0.1287	0.0307
West Bengal	3474	0.0343	0.0193	0.1222	0.02	0.1101	0.0199	-0.0144	0.006	0.0427	0.0193

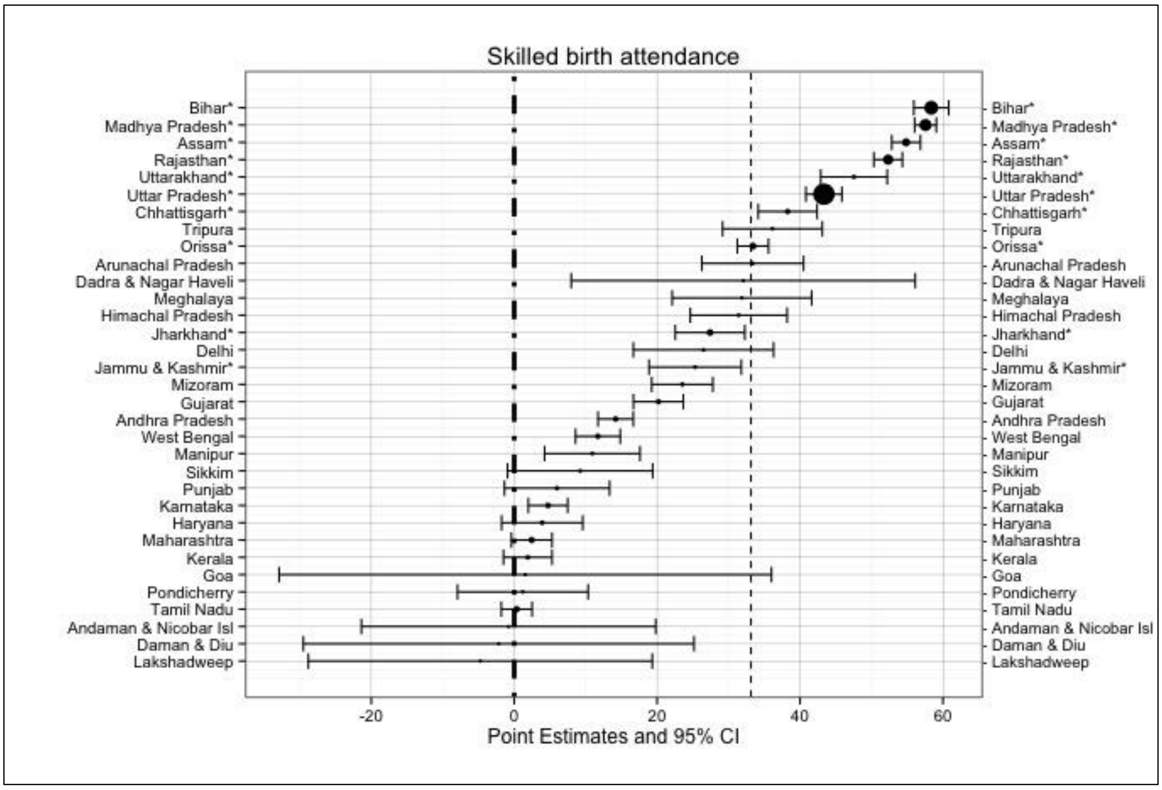
Note: NA indicates the state had no variation in outcome; N=number of observations, ANC=antenatal care, SE=standard error, IFB=in-facility birth, SBA=skilled birth attendance, PNM=perinatal morality, NNM=neonatal mortality

Appendix Figure 4: Point estimates (%) and 95 per cent CIs for state random effects relative to overall JSY mean on antenatal care.



Note: Size of point estimate relative to size of state. Vertical dashed line is pooled average. Thick vertical dot-dash line is 0. High-focus states are indicated with asterisks.

Appendix Figure 5: Point estimates (%) and 95 per cent CIs for state random effects relative to overall JSY mean on skilled birth attendance



Note: Size of point estimate relative to size of state. Vertical dashed line is pooled average. Thick vertical dot-dash line is 0. High-focus states indicated with asterisks.

References

Bafumi, J and Gelman, A, 2006. Fitting Multilevel Models When Predictors and Group Effects Correlate. Prepared for the Annual Meeting of the Midwest Political Science Association, Chicago, IL.

Bhutta, ZA, Chopra, M, Axelson, H, Berman, P, Boerma, T, Bryce, J, Bustreo, F, Cavagnero, E, Cometto, G, Daelmans, B, De Francisco, A, Fogstad, H, Gupta, N, Laski, L, Lawn, J, Maliqi, B, Mason, E, Pitt, C, Requejo, J, Starrs, A, Victora, CG and Wardlaw, T, 2010. Countdown to 2015 decade report (2000–10): taking stock of maternal, new-born, and child survival. *The Lancet,* 375(9743), pp. 2,032–2,044.

Bhutta, ZA, Das, JK, Bahl, R, Lawn, JE, Salam, RA, Paul, VK, Sankar, MJ, Blencowe, H, Rizvi, A, Chou, VB and Walker, N, 2014. Can available interventions end preventable deaths in mothers, new-born babies, and stillbirths, and at what cost? *The Lancet,* 384(9940), pp. 347–370.

Black, RE, Cousens, S, Johnson, HL, Lawn, JE, Rudan, I, Bassani, DG, Jha, P, Campbell, H, Walker, CF, Cibulskis, R, Eisele, T, Liu, L and Mathers, C, 2010. Global, regional, and national causes of child mortality in 2008: a systematic analysis. *The Lancet,* 375(9730), pp. 1,969–1,987.

Carvalho, N and Rokicki, S, 2013. Reassessment of the impact of India's Janani Suraksha Yojana conditional cash transfer program: A Replication Plan for Lim *et al.* (2010). Available at: http://www.3ieimpact.org/media/filer_public/2013/12/26/rokicki_carvalho_replication_plan.pdf [Accessed12 March 2015].

Choudhury, M and Amar Nath, HK, 2012. *An Estimate of Public Expenditure on Health in India*. New Delhi: National Institute of Public Finance and Policy.

Clark, T and Linzer, D, 2015. Should I Use Fixed Or Random Effects? *Political Science Research and Methods*, 3(2), pp. 399–408

Das, A, Rao, D and Hagopian, A, 2011. India's Janani Suraksha Yojana: further review needed. *The Lancet*, 377(9762), pp. 295–296.

Das, J and Hammer, J, 2014. Are institutional births institutionalizing deaths? World Bank Blog. Available at: http://blogs.worldbank.org/futuredevelopment/are-institutional-births-institutionalizing-deaths [Accessed 3 December 2014].

Devadasan, N, Elias, MA, John, D, Grahacharya, S and Ralte, L, 2008. A conditional cash assistance programme for promoting institutional deliveries among the poor in India: process evaluation results. *Studies in HSO&P*, 24, pp. 257–273.

DuGoff, EH, Schuler, M and Stuart, EA, 2014. Generalizing Observational Study Results: Applying Propensity Score Methods to Complex Surveys. *Health Services Research*, 49(1), pp. 284–303.

Gandhi, A, Kumar, C, Saha, P, Sahoo, BK and Sharma, A, 2011. *India Human Development Report 2011: Towards Social Inclusion*. New Delhi: Oxford University Press.

Gelman, A, 2007. Struggles with survey weighting and regression modelling. *Statistical Science,* 22(2), pp. 153–164.

Gelman, A, Carlin, JB, Stern, HS and Rubin, DR, 2004. *Bayesian Data Analysis*. 2nd ed. Boca Raton: Chapman and Hall.

Gelman, A and Hill, J, 2006. *Data analysis using regression and multilevel/hierarchical models*. New York: Cambridge University Press.

Glassman, A, Duran, D, Fleisher, L, Singer, D, Sturke, R, Angeles, G, Charles, J, Emrey, B, Gleason, J, Mwebsa, W, Saldana, K, Yarrow, K and Koblinksy, M, 2013. Impact of Conditional Cash Transfers on Maternal and New-born Health. *Journal of Health, Population and Nutrition*, 31(4, suppl. 2), pp. S48–S66.

Goldstein, H, 1995. *Multilevel Statistical Models*. 2nd ed. London: Edward Arnold.

Greene, WH, 2008. *Econometric Analysis*. 6th ed. Upper Saddle River: Prentice Hall.

Morgan, L, Beith, A and Eichler, R, 2011. *Performance-Based Incentives for Maternal Health: Taking Stock of Current Programs and Future Potentials: Health Systems, 20/20 Project.* Bethesda, MD: Abt Associates Inc.

Hunter, JE and Schmidt, FL, 2000. Fixed Effects vs. Random Effects Meta-Analysis Models: Implications for Cumulative Research Knowledge. *International Journal of Selection and Assessment,* 8(4), pp. 275–292.

International Institute for Population Sciences, 2010. *District Level Household and Facility Survey (DLHS-3), 2007–08.* Mumbai: International Institute for Population Sciences.

Government of India: Central Bureau of Health Intelligence, Directorate General of Health Services. Ministry of Health and Family Welfare, 2006. *National Health Profile of India 2006: Health Infrastructure*. Okhla Indl. Area: I G Printers Pvt. Ltd.

Lagarde, M, Haines, A and Palmer, N, 2009. The impact of conditional cash transfers on health outcomes and use of health services in low and middle income countries. *Cochrane Database of Systematic Reviews*, 4(CD008137).

Lim, SS, Dandona, L, Hoisington, JA, James, SL, Hogan, MC and Gakidou, E, 2010. India's Janani Suraksha Yojana, a conditional cash transfer programme to increase births in health facilities: An impact evaluation. *The Lancet,* 375(9730), pp. 2,009–2,023.

Lim, SS, Dandona, L, Hoisington, JA, Hogan, MC and Gakidou, E, 2011. Author's reply: India's Janani Suraksha Yojana, a conditional cash transfer programme to increase births in health facilities: an impact evaluation. *The Lancet,* 377(9762), pp. 296–297.

Lozano, R, Wang, H, Foreman, KJ, Rajaratnam, JK, Naghavi, M, Marcus, JR, Dwyer-Lindgren, L, Lofgren, KT, Phillips, D, Atkinson, C, Lopez, AD and Murray, CJ, 2011. Progress towards Millennium Development Goals 4 and 5 on maternal and child mortality: an updated systematic analysis. *The Lancet*, 378(9797), pp. 1,139–1,165.

Mazumdar, S, Mills, A and Powell-Jackson, T, 2011. Financial incentives in health: New evidence from India's Janani Suraksha Yojana. Working paper. London: Oxford University.

Ministry of Health and Family Welfare (MHFW), 2006. Janani Suraksha Yojana: Features & frequently asked questions and answers. Government of India. Available at: http://mohfw.nic.in/WriteReadData/l892s/file28-99526408.pdf [Accessed on 12 June 2015].

MHFW, 2011. Annual Report to the People on Health. Government of India. Available at: http://mohfw.nic.in/WriteReadData/l892s/6960144509Annual%20Report%20to%20the%20People%20on%20Health.pdf [Accessed on 16 June 2015].

MHFWa, n.d. Executive summary. Government of India. Available at: http://mohfw.nic.in/WriteReadData/l892s/45478965412365412.pdf [Accessed on 12 June 2015].

MHFWb, n.d. About Accredited Social Health Activist (ASHA). Government of India National Health Mission. Available at: http://nrhm.gov.in/communitisation/asha/about-asha.html [Accessed on 12 June 2015].

Mohanan, M, Bauhoff, S, La Forgia, G, Babiarz, KS, Singh, K and Miller, G, 2014. Effect of Chiranjeevi Yojana on institutional deliveries and neonatal and maternal outcomes in Gujarat, India: a difference-in-differences analysis. *Bulletin of the World Health Organization*, 92(3), pp. 187–194.

Paul, VK, 2010. India: conditional cash transfers for in-facility deliveries. *The Lancet,* 375(9730), pp. 1,943–1,944.

Randive, B, Diwan, V and De Costa, A, 2013. India's conditional cash transfer programme (the JSY) to promote institutional birth: Is there an association between institutional birth proportion and maternal mortality? *PLOS ONE*, 8(6), e67452.

Rosenbaum, PR and Rubin, DB, 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika,* 70(1), pp. 41–55.

Satapathy, DM, 2009. Janani Suraksha Yojana and 'At Birth' immunization: A study in a tertiary level health centre. *Indian Journal of Community Medicine,* 34(4), pp. 351–353.

Sharma, MP, Soni, SC, Bhattacharya, M, Datta, U, Gupta, S and Nandan, D, 2009. An assessment of institutional deliveries under JSY at different levels of health care in Jaipur District, Rajasthan. *Indian Journal of Public Health,* 53(3), pp. 177–182.

United Nations Population Fund (UNFPA) – India, 2009. *Concurrent Assessment of Janani Suraksha Yojana (JSY) in Selected States: Bihar, Madhya Pradesh, Orissa, Rajasthan, Uttar Pradesh*. New Delhi: UNFPA.

Western, B, 1998. Causal heterogeneity in comparative research: A Bayesian hierarchical modelling approach. *American Journal of Political Science*, 42(4), pp. 1,233–1,259.

World Bank, 2012. *World Development Report 2012: Gender Equality and Development*, Washington, DC: The World Bank.

Publications in the 3ie Replication Paper Series

The following papers are available from http://www.3ieimpact.org/en/publications/3ie-replication-paper-series/:

Quality evidence for policymaking: I'll believe it when I see the replication, 3ie Replication Paper 1. Brown, AN, Cameron, DB and Wood, BDK (2014)

TV, female empowerment and demographic change in rural India, 3ie Replication Paper 2. Iversen, V and Palmer-Jones, R (2014)

Reanalysis of health and educational impacts of a school-based deworming program in western Kenya Part 1: A pure replication, 3ie Replication Paper 3, part 1. Aiken, AM, Davey, C, Hargreaves, JR and Hayes, RJ (2014)

Reanalysis of health and educational impacts of a school-based deworming program in western Kenya Part 2: Alternative analyses, 3ie Replication Paper 3, part 2. Aiken, AM, Davey, C, Hayes, RJ and Hargreaves, JR (2014)

The long and short of returns to public investments in fifteen Ethiopian villages, 3ie Replication Paper 4. Bowser, WH (2015)

Recalling extra data: A replication study of Finding missing markets, 3ie Replication Paper 5. Wood, BDK and Dong, M (2015)

Replication Paper Series

International Initiative for Impact Evaluation
1625 Massachusetts Ave., NW
Suite 450
Washington, DC 20036
USA

replication@3ieimpact.org
Tel: +1 202 629 3939

www.3ieimpact.org