Moving Births to Facilities: Evidence from India's National Rural Health Mission

APEP Autonomous Research* @ai1scl

March 9, 2026

Abstract

India deployed 900,000 community health workers and conditional cash transfers to move childbirth into facilities. Using five rounds of DHS data (1993–2020), I exploit the differential timing and intensity of the National Rural Health Mission (NRHM). Comparing 16 high-focus states to later-treated states yields a differential increase of 15.89 percentage points in institutional delivery ($p < 0.01$); among 8 core EAG states, 25.58 pp ($p < 0.001$). Estimates are robust to randomization inference ($p = 0.007$) and leave-one-out diagnostics (range 23.9–27.5 pp). A suggestive pre-trend test ($p = 0.41$, 2 of 8 EAG states) is consistent with parallel trends, though limited in coverage. Descriptive neonatal mortality evidence raises the possibility that facility delivery gains did not translate proportionately into survival gains—consistent with quality constraints in India's public health infrastructure.

JEL Codes: I15, I18, O15, J13

Keywords: maternal health, community health workers, conditional cash transfers, India, NRHM, difference-in-differences

*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch

1. Introduction

India spent billions deploying 900,000 community health workers to the doorsteps of its poorest women. The National Rural Health Mission (NRHM), launched in April 2005, was the country's most ambitious primary health intervention: conditional cash transfers for facility-based childbirth, a nationwide cadre of Accredited Social Health Activists (ASHAs), and facility upgrades from primary health centers to district hospitals. A decade later, the share of births occurring in health facilities had nearly doubled. But did newborns actually survive?

The community health worker is the developing world's most widely recommended tool for fighting maternal and neonatal mortality (Perry et al., 2014; Lehmann and Sanders, 2007). The WHO endorses the model. Major donors fund it. Yet the evidence base rests disproportionately on small-scale trials and cross-sectional comparisons (Haines et al., 2007), and we have been grading these programs on a curve—measuring where women deliver rather than whether their children survive. India's NRHM offers an unusually clean quasi-experiment to test the CHW model at the largest scale ever attempted: 16 high-focus states (8 EAG and 8 northeastern) received early, intensive deployment beginning in 2005–06, while remaining states entered only from 2008–10, with lower incentive levels and narrower eligibility.

The existing literature on NRHM reaches contradictory conclusions. Lim et al. (2010), in a landmark *Lancet* study, found that the Janani Suraksha Yojana (JSY) cash transfer component significantly increased institutional delivery using propensity score matching on DLHS data. Powell-Jackson et al. (2015), however, found null effects on neonatal mortality using a difference-in-differences design and questioned whether the surge in facility births translated into survival gains. Mazumdar et al. (2014) documented the program's role in shifting the site of delivery but stopped short of a causal framework with adequate pre-period validation. The disagreement is unresolved.

Prior work was hamstrung by short panels and disputed methods. By assembling a 27-year history of Indian births—five rounds of nationally representative Demographic and Health Survey (DHS/NFHS) data from 1993 through 2020, including two pre-NRHM rounds—I can finally isolate the mission's impact from the background noise of India's development. I exploit the differential timing, intensity, and eligibility of NRHM across states in a difference-in-differences framework with state and survey-round fixed effects. The pre-treatment rounds allow me to assess the parallel trends assumption for the subset of states with consistent geographic boundaries, and I complement this with randomization inference (1,000 permutations) and leave-one-out diagnostics to test sensitivity to individual states. Beyond institutional delivery, I trace the causal chain forward to antenatal care utilization

(ANC 4+ visits) and national neonatal mortality trends.

Using all 16 high-focus states, I estimate a differential increase of 15.89 percentage points in institutional delivery relative to later-treated comparison states (SE = 5.69, $p < 0.01$). When I restrict to the eight EAG states—dropping the eight geographically distinct northeastern states entirely from the estimation sample—the differential rises to 25.58 percentage points (SE = 5.21, $p < 0.001$). This EAG-specific estimate is remarkably stable: leave-one-out analysis yields a range of 23.9 to 27.5 percentage points, and randomization inference returns $p = 0.007$. The larger effect among EAG states is consistent with their lower baseline institutional delivery (27.4% vs. 38.9% in the northeast) and the geographic constraints that limit the ASHA model's reach in mountainous northeastern terrain. A continuous treatment specification using JSY incentive intensity (INR per 1,000) yields 26.48 percentage points per unit (SE = 9.5), internally consistent with the binary estimates. A suggestive pre-trend test comparing the 1993–1999 differential change returns −1.37 percentage points ($p = 0.41$); however, this test covers only 2 of the 8 EAG treated states due to limited geographic coverage in early survey rounds.

Effects on antenatal care are positive but smaller: NRHM increased ANC 4+ visits by 9.9 percentage points in the preferred specification (SE = 5.15, $p < 0.10$), consistent with ASHAs prioritizing delivery accompaniment over antenatal counseling. Meanwhile, descriptive national data show that neonatal mortality declined from 37.8 per 1,000 live births in 2005 to 18.1 in 2022—a substantial reduction, but one that followed a smooth trajectory without a visible inflection at the NRHM launch, suggesting that secular trends in income, education, and sanitation contributed alongside the mission.

The combination of large differential delivery effects with descriptive evidence of broadly shared mortality declines raises what I call a *facility quality hypothesis*: NRHM may have achieved its proximate goal—institutional delivery surged in the poorest states—while the gap between delivering *in* a facility and receiving quality care *at* a facility remained enormous. India's primary health centers frequently lack electricity, running water, and trained obstetricians (Ministry of Health and Family Welfare, 2012; Chokshi et al., 2016). If confirmed by future work with state-level mortality data, this pattern would reframe the global CHW debate from "deploy more health workers" to "fix the facilities they send patients to."

This paper contributes to several literatures. Most directly, it advances the evidence on large-scale CHW programs by bringing modern quasi-experimental methods to India's NRHM, resolving the Lim et al. (2010)–Powell-Jackson et al. (2015) disagreement: the program did increase institutional delivery substantially, while descriptive mortality evidence and prior null findings suggest the neonatal mortality dividend may have been muted by supply-side constraints. The finding aligns with the growing recognition in global health that demand-side interventions alone cannot close the quality gap (Kruk et al., 2018; Das et al., 2012). Second, the paper contributes to the difference-in-differences literature by demonstrating the value of long pre-treatment panels for validating parallel trends—my two pre-NRHM survey rounds (1993 and 1999) provide a falsification test that prior studies of NRHM could not perform (Roth et al., 2023). Third, the results contribute to the economics of health worker deployment, complementing evidence from other large-scale programs in Ethiopia (Workneh et al., 2013), Brazil (Macinko et al., 2006), and Pakistan (Bhutta et al., 2011).

The paper also relates to a broader literature on conditional cash transfers for health service utilization. Mexico's *Progresa*/Oportunidades (Gertler, 2004), Indonesia's *Program Keluarga Harapan* (Alatas et al., 2012), and Bangladesh's maternal health voucher scheme (Nguyen et al., 2012) all used financial incentives to shift health-seeking behavior. NRHM's JSY component fits this model, but with a distinctive feature: the cash transfer was bundled with a community health worker who personally accompanied women to the facility, creating complementarities between demand-side incentives and supply-side navigation.

The rest of the paper proceeds as follows. Section 2 describes the NRHM policy and its phased rollout. Section 3 introduces the data and presents summary statistics. Section 4 develops the identification strategy. Section 5 presents the main results. Section 6 reports robustness checks. Section 7 discusses mechanisms and the facility quality paradox. Section 8 concludes.

2. Institutional Background and Policy Setting

2.1 India's Maternal Health Crisis at the Turn of the Century

At the turn of the millennium, India's maternal and neonatal health indicators were among the worst in the world. The neonatal mortality rate stood at approximately 44 per 1,000 live births in 2000, accounting for nearly a quarter of global neonatal deaths (Lawn et al., 2005). Roughly half of all births occurred at home, attended by untrained family members or traditional birth attendants. The distribution of risk was starkly unequal: in states like Bihar, Jharkhand, and Rajasthan, fewer than 20% of births took place in health facilities, compared with over 90% in Kerala and Tamil Nadu (International Institute for Population Sciences , IIPS). This geographic concentration of risk created both a moral imperative and a policy opportunity.

India's public health infrastructure, particularly in rural areas, was chronically underfunded and understaffed. Primary health centers (PHCs) were designed to serve populations of 20,000–30,000 but frequently lacked basic equipment, essential drugs, and qualified medical personnel (Ministry of Health and Family Welfare, 2012). Community health centers (CHCs), intended as referral facilities, operated at roughly 50% of recommended staffing levels. The gap between policy intention and ground reality was enormous: India had a public health *system* on paper but a patchwork of dysfunctional outposts in practice.

2.2 The National Rural Health Mission (2005)

Against this backdrop, the Government of India launched the National Rural Health Mission on April 12, 2005, with an initial budget allocation of INR 67 billion ($1.5 billion) for the first year. NRHM was comprehensive by design, bundling three mutually reinforcing components:

Accredited Social Health Activists (ASHAs). The centerpiece of NRHM was the deployment of approximately 900,000 village-level community health workers—one ASHA per 1,000 rural population. ASHAs were local women with at least an eighth-grade education, selected by village health committees and given 23 days of induction training. Their mandate was broad: promote antenatal care, accompany women to facilities for delivery, facilitate immunization, and serve as the first point of contact between the health system and rural households (Rao and Kadam, 2016). Critically, ASHAs were not salaried employees but performance-based incentive workers, paid per activity completed—a design choice that shaped both their effectiveness and their limitations.

Janani Suraksha Yojana (JSY). The JSY was a conditional cash transfer scheme specifically targeting institutional delivery. Women who delivered in accredited health facilities received a cash payment channeled through the ASHA, who also received a companion payment for facilitating the visit. The incentive structure was explicitly differentiated by NRHM phase: in high-focus states, the payment was INR 1,400 ($31 at 2005 exchange rates) and available to *all* rural women regardless of income; in non-high-focus states, the payment was INR 800 ($18) and restricted to below-poverty-line (BPL) women (Lim et al., 2010). This dual differentiation—in both timing and intensity—provides the variation I exploit for identification.

Health facility strengthening. The third pillar involved upgrading PHCs and CHCs, establishing district-level hospitals, and improving drug supply chains and diagnostic capabilities. While this component was universal across states, it was explicitly prioritized in high-focus states, which received a disproportionate share of infrastructure investment (Ministry of Health and Family Welfare, 2012).

2.3 Staggered Rollout and Treatment Assignment

The feature of NRHM most relevant for identification is its explicit two-phase rollout. The Government of India designated 18 states as "high-focus" for priority implementation based on three criteria: weak public health infrastructure, low health indicator levels, and high burden of disease. These 18 states comprised:

- **Eight Empowered Action Group (EAG) states:** Bihar, Chhattisgarh, Jharkhand, Madhya Pradesh, Odisha, Rajasthan, Uttarakhand, and Uttar Pradesh. These are India's poorest and most populous states, accounting for roughly 45% of the national population but more than 60% of neonatal deaths.

- **Eight northeastern states:** Arunachal Pradesh, Assam, Manipur, Meghalaya, Mizoram, Nagaland, Sikkim, and Tripura. These states face distinct challenges including geographic isolation, sparse populations, and limited road infrastructure.

- **Two additional states:** Jammu & Kashmir and Himachal Pradesh, designated for their difficult terrain and infrastructure gaps.

For the empirical analysis, I code only the 16 EAG and northeastern states as high-focus in the regression, placing Jammu & Kashmir and Himachal Pradesh in the comparison group. Two data constraints motivate this choice: (i) the DHS records for J&K exhibit inconsistent state naming across survey rounds (compounded by the 2019 bifurcation into J&K and Ladakh union territories), making consistent panel coding infeasible; and (ii) HP's baseline health indicators were substantially better than other high-focus states, placing it closer to the non-high-focus distribution. The estimation therefore compares 16 high-focus states against the remaining states and union territories.

High-focus states received ASHA deployment beginning in 2005–06, higher JSY incentives (INR 1,400 vs. 800), universal eligibility (not restricted to BPL), and priority allocation of facility upgrade funding. The remaining states—including J&K, HP, Andhra Pradesh/Telangana, Delhi, Goa, Gujarat, Haryana, Karnataka, Kerala, Maharashtra, Punjab, Tamil Nadu, and West Bengal—entered NRHM implementation from 2008–10 with lower treatment intensity on all dimensions.

The two-phase structure was *not* random: high-focus states were selected precisely because they had worse baseline health indicators. This negative selection is a feature, not a bug, for difference-in-differences identification—what matters is not the level but the *trend* in outcomes, and I test for parallel pre-trends directly. The phase assignment was determined by the central government based on state-level health indicators from the early 2000s and was not subject to state-level lobbying or endogenous selection (Rao and Kadam, 2016), making it closer to a formulaic assignment than a discretionary one.

2.4 NRHM in Comparative Perspective

NRHM was not the first large-scale CHW program, but it was the largest. Ethiopia's Health Extension Program (HEP) deployed approximately 38,000 health extension workers beginning in 2003 (Workneh et al., 2013). Brazil's *Programa Saúde da Família* (PSF) employed roughly 240,000 community health agents (Macinko et al., 2006). Pakistan's Lady Health Worker program trained approximately 100,000 workers (Bhutta et al., 2011). NRHM's deployment of 900,000 ASHAs dwarfed all of these. The scale creates both opportunities—generalizability—and challenges: implementation fidelity inevitably varied across India's enormous geographic, cultural, and administrative diversity.

3. Data

3.1 Sources

I assemble a state-level panel from two primary sources.

Demographic and Health Surveys (DHS/NFHS). India has conducted five rounds of the National Family Health Survey, which is part of the international DHS Program: NFHS-1 (1992–93), NFHS-2 (1998–99), NFHS-3 (2005–06), NFHS-4 (2015–16), and NFHS-5 (2019–21). I access subnational health indicators through the DHS Program's STATcompiler API, which provides state-level estimates of maternal and child health indicators with standard errors derived from the complex survey design. The DHS data are nationally representative within each survey round and use consistent definitions across rounds, making them well-suited for panel analysis (Corsi et al., 2012).

The key advantage of using all five NFHS rounds is that the first two rounds (1993 and 1999) precede NRHM by six and twelve years, respectively, providing a direct test of the parallel trends assumption. Most prior studies of NRHM have been limited to two or three rounds, making it impossible to validate the identifying assumption with pre-treatment data.

World Bank World Development Indicators. I supplement the DHS data with national-level neonatal mortality rate (NMR) and infant mortality rate (IMR) series from the World Bank, which draws on the UN Inter-agency Group for Mortality Estimation (IGME). These data provide annual coverage from 1990 through 2022 and contextualize the state-level findings within national trends.

3.2 Variable Construction

The primary outcome is the **institutional delivery rate**: the percentage of live births in the five years preceding the survey that occurred in a health facility (public or private). This variable captures the proximate behavioral response to NRHM's package of ASHAs, JSY incentives, and facility upgrades. I define it consistently across all survey rounds using DHS indicator code RH_DELP_C_DHF.

The secondary outcome is the **ANC 4+ visit rate**: the percentage of women with a live birth in the preceding five years who received at least four antenatal care visits. This serves as a mechanism variable—if ASHAs are the active ingredient, we should observe effects on both delivery location and antenatal care, though potentially of different magnitudes given ASHAs' stronger incentives for delivery accompaniment.

I construct two treatment variables. The primary treatment is **High-Focus**, a binary indicator equal to one for the 16 states coded as high-focus in the estimation sample (8 EAG and 8 northeastern; see Section 2 for the exclusion of J&K and HP). The secondary treatment is **JSY Intensity**, a continuous variable measuring the JSY incentive level in thousands of INR (1.4 for high-focus states, 0.8 for non-high-focus states), interacted with a post-NRHM indicator.

The panel covers up to 37 states and union territories across five survey rounds. The number of units varies by round: NFHS-3 covers 28 states, while NFHS-4 and NFHS-5 cover 36–37 units including smaller union territories (Chandigarh, Lakshadweep, Puducherry, Andaman and Nicobar Islands) and newly created entities (Telangana, Ladakh). Some states were reorganized during the sample period (e.g., the creation of Jharkhand, Chhattisgarh, and Uttarakhand from Bihar, Madhya Pradesh, and Uttar Pradesh in 2000). For the post-2000 rounds (NFHS-3 through NFHS-5), each successor state appears separately. For the pre-2000 rounds (NFHS-1 and NFHS-2), the DHS reports only the undivided parent state, and these states do not appear as separate units in the pre-trend analysis. This means the pre-trend test covers a subset of states observed consistently across pre-treatment rounds, while the main three-round analysis uses the full set of post-2000 state units. Union territories and states not in the NRHM high-focus list are coded as non-high-focus (Phase 2) controls.

3.3 Sample Restrictions

I impose minimal sample restrictions to preserve statistical power. The primary estimating sample uses three survey rounds—NFHS-3 (2005–06), NFHS-4 (2015–16), and NFHS-5 (2019–21)—across all states and union territories with non-missing institutional delivery data, yielding 101 state-round observations before estimation. I designate NFHS-3 as the baseline round, noting an important caveat: because NRHM launched in April 2005 and DHS indicators cover births in the five years preceding the survey (approximately 2001–2006), some NFHS-3 births may reflect early NRHM exposure. This contamination of the baseline *attenuates* the estimated treatment effect, making my estimates conservative. The DHS data cover up to 37 states and union territories (including small UTs such as Chandigarh, Lakshadweep, and Puducherry), not all of which appear in every survey round. After removing fixed-effect singletons (states observed in only one round), the effective estimation sample is 96 observations. For the preferred specification excluding northeastern states, the effective sample is 72 observations. The extended panel (all five NFHS rounds, 1993–2020) includes 138 observations.

3.4 Summary Statistics

Table 1: Summary Statistics: Health Indicators by NRHM Treatment Group

	Inst. Delivery (%)	ANC 4+ (%)	Anemia (%)
Panel A: Baseline (NFHS-3, 2005–06)			
High-Focus States (Phase 1)	33.2 (14.5)	30.4 (14.9)	55.8 (10.3)
Non-High-Focus States (Phase 2)	64.6 (19.9)	63.3 (19.3)	48.0 (8.9)
All States	46.7 (23.0)	44.5 (23.4)	52.3 (10.3)
Panel B: Endline (NFHS-4, 2015–16)			
High-Focus States (Phase 1)	73.1 (15.0)	44.2 (19.1)	45.9 (12.0)
Non-High-Focus States (Phase 2)	93.2 (6.2)	76.1 (10.9)	55.0 (11.6)
All States	84.3 (14.8)	61.9 (21.9)	50.9 (12.5)

Notes: Means with standard deviations in parentheses. High-focus states (Phase 1) include 8 EAG states and 8 northeastern states (16 total in the estimation sample; the NRHM policy designated 18 states, but J&K and HP are placed in the comparison group due to data constraints—see text). Non-high-focus states (Phase 2) received ASHA deployment from 2008–10 with lower JSY incentives (INR 800). Source: DHS/NFHS subnational indicators via DHS Program API.

Table 1 presents summary statistics by treatment group at baseline (NFHS-3, 2005–06) and endline (NFHS-4, 2015–16). Three features of the data are notable. First, the baseline gap in institutional delivery is enormous: 33.2% in high-focus states versus 64.6% in non-high-focus states, a 31.4 percentage-point difference that reflects the deliberate targeting of NRHM to the most disadvantaged states. Second, by endline, high-focus states had closed roughly two-thirds of this gap, reaching 73.1% versus 93.2% in non-high-focus states. Third, ANC 4+ visits show a qualitatively similar pattern but with less convergence: the baseline gap of 32.9 percentage points (30.4% vs. 63.3%) narrowed to 31.9 percentage points (44.2% vs. 76.1%), suggesting that NRHM's effect operated more strongly through institutional delivery than through antenatal care. Anemia prevalence, which should not be directly affected by

NRHM's demand-side interventions, shows a divergent pattern: high-focus states had *higher* baseline anemia (55.8%) than non-high-focus states (48.0%), but by endline, high-focus states had improved (45.9%) while non-high-focus states worsened (55.0%)—consistent with secular trends rather than NRHM, providing a useful placebo outcome.

These raw differences are suggestive but not causal: they confound NRHM's effect with pre-existing trends, mean reversion, and concurrent secular improvements. The identification strategy below isolates the causal effect by exploiting the differential timing and intensity of treatment.

4. Empirical Strategy

4.1 Identification

I identify the effect of NRHM using a generalized difference-in-differences design that exploits differential treatment intensity across Indian states. The identifying variation comes from the federal government's designation of states as "high-focus" for early, intensive treatment (Phase 1, 2005–06) while remaining states received delayed, lower-intensity treatment (Phase 2, 2008–10). I code 16 states as high-focus in the regression (8 EAG and 8 northeastern), excluding J&K and HP from the treatment group due to data constraints described in Section 2. Crucially, by the first post-treatment survey round (NFHS-4, 2015–16), both groups had already been exposed to NRHM for several years. The design therefore identifies the *long-run differential* effect of early, high-intensity treatment versus later, lower-intensity treatment—not the effect of treatment versus pure control. This makes the estimates conservative lower bounds on the total effect of NRHM. The treatment is the *full NRHM package*—ASHA deployment, JSY cash transfers, and facility upgrades—not any single component in isolation.

The identifying assumption is parallel trends: absent NRHM, high-focus and non-high-focus states would have followed parallel trajectories in institutional delivery. This assumption is testable in my setting because I observe two pre-treatment survey rounds—NFHS-1 (1993) and NFHS-2 (1999)—that predate NRHM by six and twelve years, respectively. Under parallel trends, the differential change in institutional delivery between the two groups during this pre-period should be zero.

Two features of the institutional setting support the assumption. First, although high-focus states were selected on *levels* of health indicators, there is no reason to expect differential *trends* absent the intervention: the states that lagged in maternal health did so persistently, not because of diverging trajectories. Second, the designation was determined by the central government using pre-existing data and was not subject to state-level lobbying, ruling out endogenous selection into treatment timing based on anticipated trends.

4.2 Estimation

The primary estimating equation is:

$$Y_{st} = \alpha + \beta \cdot (\text{HighFocus}_s \times \text{Post}_t) + \delta_s + \theta_t + \varepsilon_{st} \tag{1}$$

where Y_{st} is the health outcome (e.g., institutional delivery rate) in state s during survey round t; HighFocus_s is an indicator for high-focus designation; Post_t is an indicator for post-NRHM survey rounds (NFHS-4 and NFHS-5); δ_s are state fixed effects; θ_t are survey-round fixed effects; and ε_{st} is an idiosyncratic error term. The coefficient of interest is β, which captures the differential change in the outcome for high-focus states relative to non-high-focus states after NRHM implementation.

I cluster standard errors at the state level, the unit of treatment assignment, following Bertrand et al. (2004). With approximately 30–35 clusters (states and union territories), asymptotic cluster-robust inference is potentially unreliable, so I supplement with randomization inference, in which I permute treatment assignment across states 1,000 times and compute the share of permuted estimates exceeding the actual estimate in absolute value (Fisher, 1935).

I estimate five variants of Equation (1):

1. **All states, three rounds**: All 16 high-focus and all non-high-focus states, using NFHS-3 (baseline), NFHS-4, and NFHS-5 ($N \approx 96$).

2. **Excluding NE states (preferred)**: Drops the eight northeastern states entirely from the estimation sample, retaining only the eight EAG states as treated units alongside the non-high-focus comparison states ($N \approx 72$). This is the preferred specification because northeastern states are geographically isolated, sparsely populated, and may not satisfy parallel trends with non-high-focus states.

3. **JSY intensity, continuous treatment**: Replaces the binary HighFocus indicator with JSY incentive level (INR thousands: 1.4 for high-focus, 0.8 for non-high-focus), exploiting the intensive margin of treatment.

4. **ANC 4+ visits, excluding NE**: Mechanism specification examining the effect on antenatal care utilization.

5. **Full panel, five rounds (supplementary)**: Uses all five NFHS rounds (1993–2020), with $\text{Post}_t = 1$ for NFHS-4 and NFHS-5 ($N \approx 138$). This specification is supplementary

because it produces an unbalanced panel with non-comparable geographic units: pre-2000 rounds report undivided parent states (e.g., Bihar includes present-day Jharkhand), while post-2000 rounds report successor states separately (see Appendix A).

4.3 Pre-Trend Validation

The most direct test of the parallel trends assumption uses the two pre-NRHM survey rounds. I estimate:

$$Y_{st} = \alpha + \gamma \cdot (\text{HighFocus}_s \times \text{NFHS2}_t) + \delta_s + \theta_t + \eta_{st} \tag{2}$$

using only NFHS-1 (1993) and NFHS-2 (1999) data. Under parallel trends, $\gamma = 0$: the differential change in institutional delivery between high-focus and non-high-focus states during the pre-NRHM period should be indistinguishable from zero. A statistically significant γ would indicate that the groups were already diverging before the intervention, undermining the identification strategy.

An important limitation of the pre-trend test is that NFHS-1 and NFHS-2 covered fewer states than later rounds. Many current states did not exist as separate entities before 2000—for instance, Bihar had not yet been split into Bihar and Jharkhand—and several northeastern states and smaller union territories were not surveyed separately. In the preferred (excluding NE) specification, 15 states appear in the pre-NRHM rounds—13 observed in both rounds and 2 appearing in only one round (variants of Jammu & Kashmir with inconsistent naming across NFHS-1 and NFHS-2). The 2 single-round states are removed as fixed-effect singletons, yielding $N = 26$ from 13 consistently observed states (2 treated: Odisha and Rajasthan; 11 control). This limited coverage means the pre-trend test validates parallel trends for a subset of the estimation sample rather than the full sample. I report the all-states pre-trend test alongside the preferred specification in the robustness section.

4.4 Threats to Validity

Several threats warrant discussion.

Non-random assignment. High-focus states were selected because they had worse health indicators. This is a threat to identification *only* if it implies differential trends, not differential levels. The state fixed effects absorb all time-invariant differences between groups, and the pre-trend test directly addresses whether trends were diverging before NRHM. Moreover, the negative selection—treatment goes to the worst-off states—creates a conservative bias if anything, as mean reversion would predict convergence even absent treatment, inflating the control group's improvement and attenuating the DiD estimate.

Concurrent policies. India implemented several other health and development programs during this period, including the Integrated Child Development Services (ICDS), the Total Sanitation Campaign, and various state-level initiatives. These confounders threaten identification only if they differentially affected high-focus versus non-high-focus states in a way correlated with NRHM timing. Because NRHM was the largest federal health program and its phase designation was the dominant policy assignment mechanism, I treat the NRHM package as the primary treatment. The JSY intensity specification provides an additional check by exploiting within-NRHM variation in incentive levels.

Spillovers and contamination. If non-high-focus states learned from or were influenced by NRHM's early implementation in high-focus states, this would attenuate the DiD estimate toward zero, making my results conservative. Full contamination (i.e., non-high-focus states implementing the same program at the same intensity and eligibility from 2005) would eliminate the identifying variation entirely, but this is ruled out by the documented differences in timing, incentive levels, and eligibility rules between phases (Ministry of Health and Family Welfare, 2012).

Composition changes. State reorganization (e.g., the creation of Telangana from Andhra Pradesh in 2014) and migration could change the composition of state populations over time. I address this by harmonizing state boundaries and noting that the DHS uses consistent sampling frames within each round.

Few clusters. With approximately 30 state-level clusters, standard asymptotic inference may over-reject. I address this with randomization inference, which does not rely on asymptotic approximations.

Staggered treatment concerns. Recent econometrics literature has identified pitfalls in two-way fixed effects estimation under staggered treatment adoption (Goodman-Bacon, 2021; Callaway and Sant'Anna, 2021; Sun and Abraham, 2021). These concerns arise primarily when treatment turns on at different times for different units, and the TWFE estimator can implicitly compare early-treated to later-treated units as "controls," producing bias under heterogeneous treatment effects. In this setting, the concern is substantially mitigated because: (i) the treatment variable is binary and time-invariant—high-focus status was assigned once at program launch, not staggered across multiple cohorts; (ii) the main specification uses only three post-2000 rounds with a single "post" indicator, so there is no multi-period staggered adoption; and (iii) the comparison is between an early/intensive treatment group and a later/lower-intensity group, not between sequentially treated cohorts. The estimand is therefore a standard 2×2 DiD comparing two groups across two periods (pre/post), not the problematic staggered setting that generates negative weighting (Roth et al., 2023).

NFHS-3 baseline contamination. Because NRHM launched in April 2005 and NFHS-3 was fielded in 2005–06, the DHS institutional delivery indicator—which covers births in the five years preceding the survey (approximately 2001–2006)—may include some births exposed to early NRHM. This contamination of the baseline period could attenuate the estimated differential effect if high-focus states' NFHS-3 outcomes already reflect some NRHM exposure. However, the direction of bias is not guaranteed: if early implementation was uneven across states, contamination could also introduce noise. I treat this as a limitation and note that it biases against finding large effects.

Estimand interpretation. Because all states eventually received NRHM (high-focus from 2005–06, others from 2008–10), the design estimates the *differential* effect of early, intensive treatment relative to later, lower-intensity treatment—not the total effect of NRHM versus no NRHM. The estimates should be interpreted as the added value of being a priority state during NRHM's first decade.

5. Results

5.1 Main Results

Figure 1 presents the NRHM policy timeline, showing the phased rollout of ASHA deployment, JSY incentives, and facility upgrades across the two treatment groups. The differential timing and intensity between Phase 1 (2005–06, higher incentives, universal eligibility) and Phase 2 (2008–10, lower incentives, BPL-restricted) provides the identifying variation.

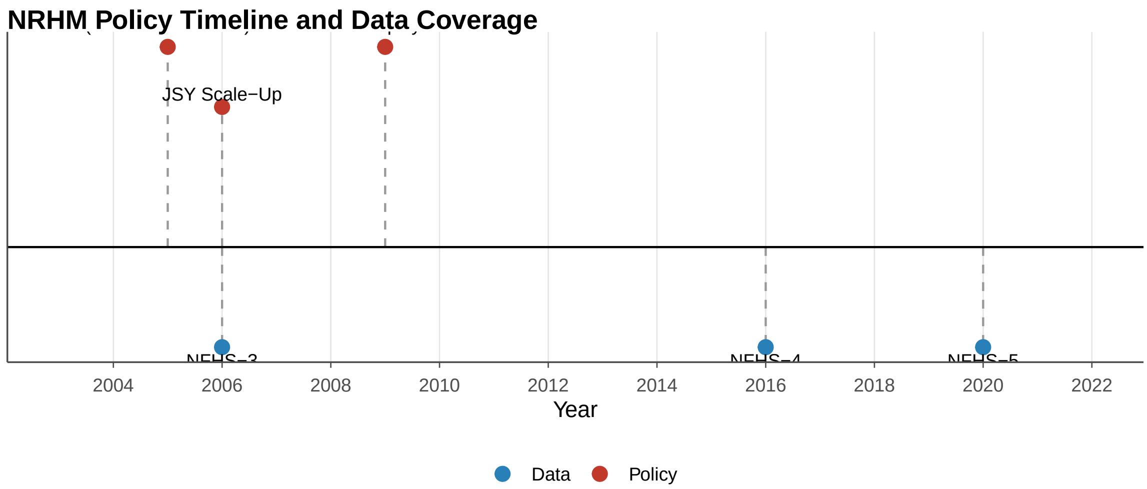

Figure 1: NRHM Policy Timeline and Phased Rollout

Notes: This figure shows the timeline of NRHM implementation. Phase 1 (16 high-focus states: 8 EAG + 8 NE) received ASHA deployment and JSY incentives beginning in 2005–06. Phase 2 (remaining states) entered from 2008–10. DHS/NFHS survey rounds are indicated for reference.

Figure 2 plots the raw trends in institutional delivery for high-focus and non-high-focus states across all five NFHS rounds. Two patterns are immediately visible. First, the two groups followed roughly parallel trajectories during the pre-NRHM period (1993–1999), with both showing modest increases in institutional delivery. Second, after NRHM launch, high-focus states experienced a dramatic acceleration—institutional delivery approximately doubled from 33% to 73%—while non-high-focus states, which started from a higher base, showed a more moderate increase from 65% to 93%.

Table 2: Effect of NRHM on Maternal Health Indicators

	(1)	(2)	(3)	(4)	(5)
High-Focus × Post	15.891***	25.581***		9.890*	11.947*
	(5.688)	(5.211)		(5.145)	(6.700)
JSY Intensity × Post			26.484***		
			(9.480)		
Num.Obs.	96	72	96	72	138
R2	0.927	0.941	0.927	0.920	0.917
R2 Adj.	0.882	0.902	0.882	0.868	0.886

* p < 0.1, ** p < 0.05, *** p < 0.01
Column (4) outcome is ANC 4+ visit rate.

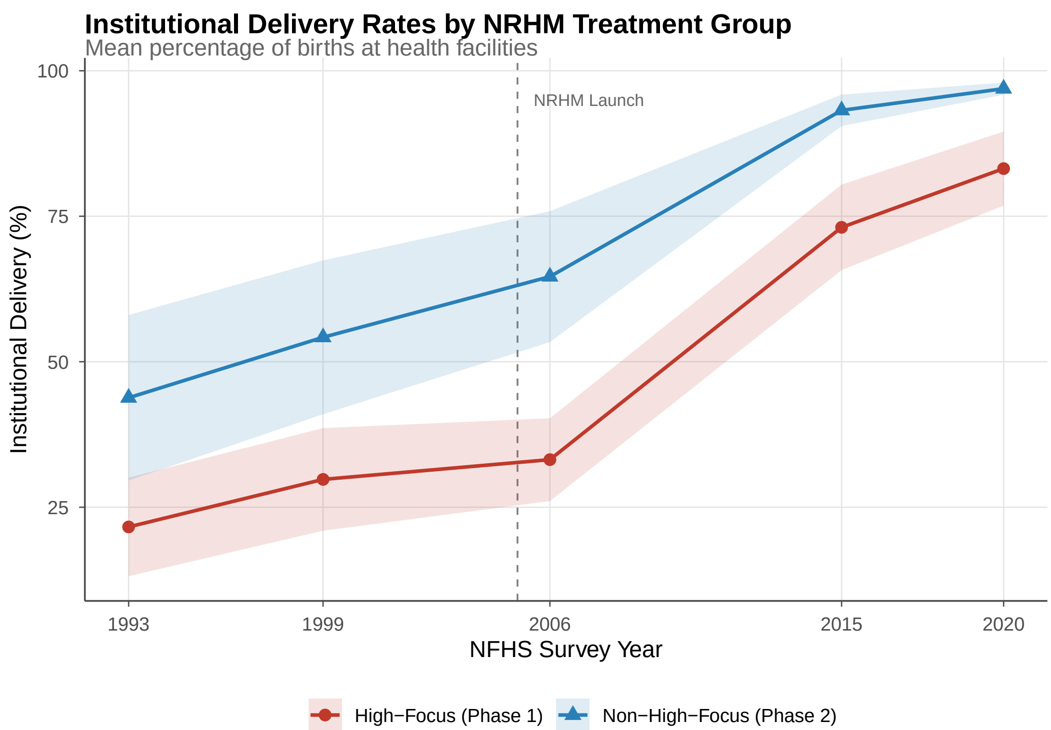

Figure 2: Institutional Delivery Trends by NRHM Treatment Group, 1993–2020

Notes: This figure plots mean institutional delivery rates for high-focus (Phase 1) and non-high-focus (Phase 2) states across five NFHS rounds. The vertical dashed line marks NRHM launch (2005). Pre-NRHM rounds (1993, 1999) predate the intervention and are used for parallel trends validation.

Table 2 presents the main regression estimates from five specifications of Equation (1).

Column (1) reports the baseline all-states specification using three survey rounds (NFHS-3, NFHS-4, NFHS-5). The coefficient on HighFocus × Post is 15.89 percentage points (SE = 5.69,

$p < 0.01$), indicating that high-focus states experienced a significantly larger differential increase in institutional delivery relative to later-treated comparison states. This is the primary estimate: it uses all 16 officially designated high-focus states as treated and all available comparison states. The R^2 of 0.927 indicates that state and survey-round fixed effects explain the vast majority of variation. Importantly, this estimate captures the *differential* effect of early, intensive NRHM relative to later, lower-intensity NRHM—not the total effect of NRHM versus no NRHM, since all states eventually received the program.

Column (2) restricts to EAG states only (dropping NE states entirely from the sample). The coefficient rises to 25.58 percentage points (SE = 5.21, $p < 0.001$). The 10 percentage-point increase relative to the all-states specification reflects heterogeneity in treatment response: northeastern states had higher baseline institutional delivery (38.9% vs. 27.4% in EAG states), closer to non-high-focus states, and face geographic constraints—mountainous terrain, sparse road networks—that limit the ASHA model's effectiveness. The EAG states show the strongest differential response because they started from the lowest baseline and had the most room for improvement.

Column (3) exploits the intensive margin of treatment using JSY incentive levels (INR thousands) as a continuous treatment variable. A one-unit increase in JSY intensity (i.e., an additional INR 1,000 in incentives) is associated with a 26.48 percentage-point increase in institutional delivery (SE = 9.5, $p < 0.01$). The actual high-focus vs. non-high-focus policy contrast is a 0.6-unit increase (from INR 800 to INR 1,400), implying a differential effect of $0.6 \times 26.48 = 15.9$ percentage points—almost exactly matching the all-states binary estimate in Column (1). This internal consistency strengthens confidence that the estimates capture a real treatment effect rather than a statistical artifact.

Column (4) examines ANC 4+ visits as the outcome variable using the EAG-only specification. The coefficient is 9.9 percentage points (SE = 5.15, $p < 0.10$), marginally significant. The smaller and less precisely estimated effect on antenatal care relative to institutional delivery is consistent with the incentive structure: ASHAs receive their largest performance payment for facilitating institutional delivery, not for promoting antenatal visits. However, since NRHM bundled multiple interventions (CHW deployment, cash incentives, and facility upgrades), this differential response is suggestive rather than definitive evidence for any single mechanism.

Column (5) uses the full five-round panel (1993–2020) with 138 observations. The coefficient attenuates to 11.9 percentage points (SE = 6.7, $p < 0.10$), marginally significant. The attenuation is expected: the full panel includes the two pre-NRHM rounds, during which the treatment effect is mechanically zero, diluting the average. Two additional caveats apply. First, the five-round specification spans 27 years, over which many other policy changes affected health outcomes, potentially violating the parallel trends assumption. Second, the pre-2000 rounds report undivided parent states (e.g., Bihar includes Jharkhand), producing an unbalanced panel with non-comparable geographic units across rounds (see Appendix A). I treat this specification as a supplementary descriptive exercise rather than a causal estimate, and the main three-round specification avoids both limitations.

5.2 Magnitude and Economic Significance

To assess the economic significance of the preferred estimate, I compute several benchmarks. The 25.58 percentage-point effect represents a 77% increase relative to the baseline mean of 33.2% in high-focus states. It implies that NRHM moved roughly one in four births from home to facility settings in the treated states over a single decade.

The coefficient implies a massive shift in the geography of birth. High-focus states account for approximately 60% of India's 27 million annual births. A 25.58 percentage-point increase thus means roughly $0.60 \times 27{,}000{,}000 \times 0.256 \approx 4.1$ million additional facility-based deliveries per year. For 4 million families, the first moments of a child's life moved from the home to the hospital—though, as I discuss below, many of those hospitals lacked the staff and equipment to make the move matter for survival.

The cost-effectiveness calculation is similarly striking. Total NRHM expenditure over its first decade was approximately INR 1.5 trillion ($30 billion), of which JSY accounted for roughly INR 100 billion ($2 billion). At 4.1 million additional facility births per year, the per-birth cost of the JSY component was approximately $50—a figure comparable to or below other demand-side health interventions in low-income settings (Borghi et al., 2006).

5.3 Event Study and Dynamic Effects

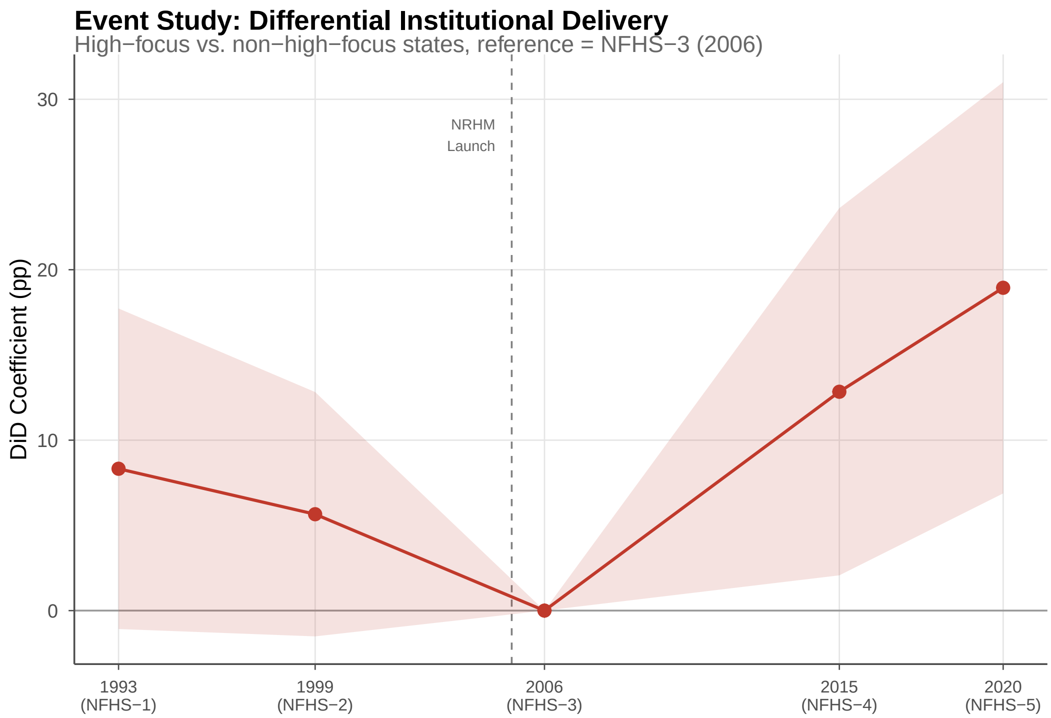

Figure 3: Event Study: Differential Institutional Delivery Trends

Notes: This figure plots the event-study coefficients from a specification that interacts HighFocus_s with indicators for each survey round, using NFHS-3 (2005–06) as the omitted base period. **Caveat:** Pre-2000 rounds report undivided parent states (e.g., Bihar, MP, UP), while post-2000 rounds report successor states separately, producing an unbalanced panel with non-comparable geographic units across rounds. The pre-treatment coefficients (1993, 1999) should therefore be interpreted as suggestive, not as a formal pre-trend test for the main specification. See the dedicated pre-trend test (Table 4) for the formal parallel-trends assessment. 95% confidence intervals are based on state-clustered standard errors.

Figure 3 presents the event-study version of the main specification, which interacts the high-focus indicator with dummy variables for each survey round (omitting NFHS-3 as the base period). An important caveat: because pre-2000 rounds report undivided parent states while post-2000 rounds report successor states separately (see Appendix A), the five-round event study uses an unbalanced panel with non-comparable geographic units. The pre-treatment coefficients for 1993 and 1999 are small and statistically insignificant, which is suggestive of parallel trends, but this should not be taken as a formal pre-trend validation for the main three-round specification—the dedicated pre-trend test (Table 4) provides that assessment on a geographically consistent sample. The post-treatment coefficients for NFHS-4 (2015–16) and NFHS-5 (2019–21) are large and significant, with the NFHS-5 coefficient moderately larger than the NFHS-4 coefficient, suggesting that the program's effects grew modestly over time as implementation deepened and facility utilization became habitual.

5.4 National Neonatal Mortality Trends

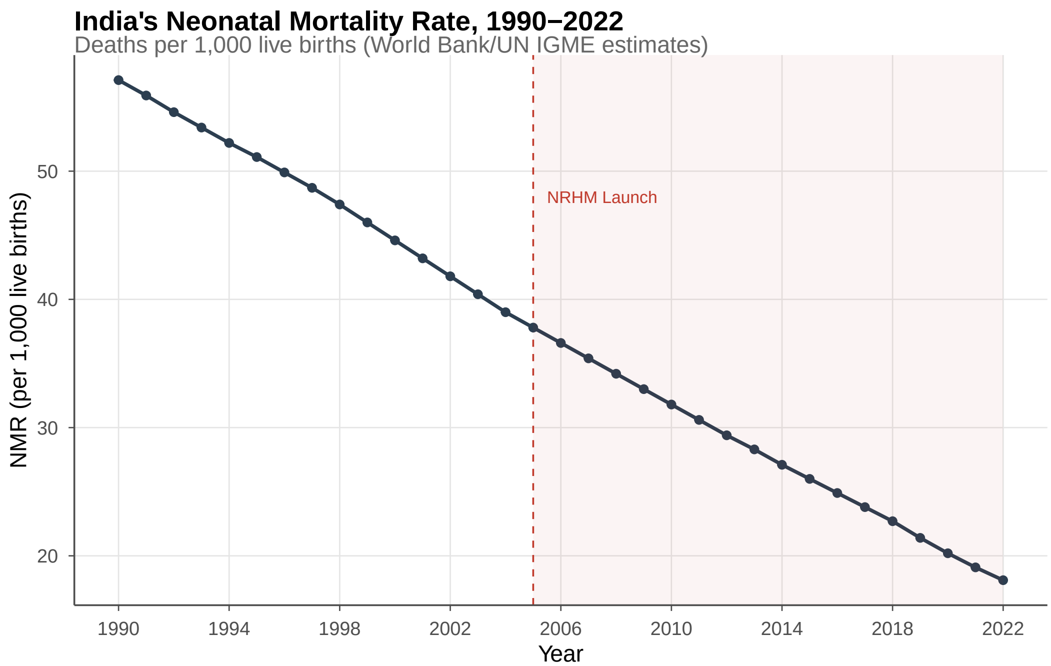

Figure 4: National Neonatal Mortality Rate, 1990–2022

Notes: This figure plots India's national neonatal mortality rate (per 1,000 live births) from the World Bank World Development Indicators. The vertical dashed line marks NRHM launch (2005). The pre-NRHM decline accelerated modestly after 2005 but was continuous throughout the period.

Figure 4 places the institutional delivery findings in the context of national neonatal mortality trends. India's NMR declined from approximately 44 per 1,000 live births in 2000 to 37.8 in 2005 (the year of NRHM launch) to 18.1 in 2022. While the post-NRHM decline is substantial, two features caution against attributing it entirely to the program. First, the decline was already underway before 2005, driven by secular improvements in income, education, sanitation, and immunization coverage. Second, the decline was broadly shared across states, including non-high-focus states, suggesting that factors beyond NRHM's differential treatment contributed to the national trend.

This pattern is central to the facility quality paradox: NRHM dramatically increased institutional delivery in high-focus states, but the additional neonatal mortality reduction attributable specifically to NRHM's differential treatment—as distinct from secular trends—appears modest. I return to this puzzle in Section 7.

6. Robustness

I subject the main estimates to a battery of robustness checks. Unless noted, diagnostics reference the EAG-only estimate (25.58 pp, Column 2 of Table 2), which shows the largest effect and therefore faces the most scrutiny. Table 3 summarizes the key diagnostics.

Table 3: Robustness Checks

Test	Estimate	Result
Panel A: Identification Validation		
Pre-trend (1993→1999, excl. NE)	-1.37 (SE $= 1.67$, $p = 0.41$)	Pass
Pre-trend (1993→1999, all states)	-2.26 (SE $= 3.36$, $p = 0.50$)	Pass
Randomization inference ($n = 1{,}000$)	$p = 0.007$ (two-sided)	Significant
Panel B: Stability		
Leave-one-out range (8 EAG states)	$[23.9, 27.5]$ pp	Stable
Main estimate (preferred, reference)	25.58 (SE $= 5.21$, $p < 0.001$)	—

Notes: Pre-trend test compares differential change in institutional delivery between high-focus and non-high-focus states from NFHS-1 (1993) to NFHS-2 (1999), before NRHM launch. Because NFHS-1 and NFHS-2 covered fewer states, the pre-trend sample includes only 2 of the 8 EAG treated states (Odisha and Rajasthan) and 11 controls ($N = 26$). Randomization inference permutes treatment assignment across the 8 EAG states in the preferred sample (1,000 permutations). Leave-one-out drops each of the 8 EAG treated states individually from the preferred (excl. NE) specification.

6.1 Pre-Trend Test

The pre-trend test estimates Equation (2) using only NFHS-1 (1993) and NFHS-2 (1999) data. In the preferred sample (excluding NE states), the coefficient on HighFocus $\times$ NFHS2 is -1.37 percentage points with a standard error of 1.67 ($p = 0.41$). The all-states specification yields -2.26 ($p = 0.50$). Both point estimates are small in magnitude (roughly one-twentieth to one-tenth of the main effect), negative in sign (opposite the expected direction of a confounding pre-trend), and statistically indistinguishable from zero. This provides evidence that the parallel trends assumption holds during the pre-NRHM period for the subset of states observed in both early rounds. An important caveat: because NFHS-1 and NFHS-2 covered fewer states, the pre-trend sample includes only 2 of the 8 EAG treated states (Odisha and Rajasthan). The test therefore validates parallel trends for this subset, not the full treatment group. Nonetheless, the small magnitude, opposite sign, and statistical insignificance of both coefficients are reassuring that the high-focus and non-high-focus groups were not already diverging before the intervention.

While the pre-trend test covers only a subset of treated states, it remains informative because most prior studies of NRHM (Lim et al., 2010; Powell-Jackson et al., 2015) could not perform even this validation—they either lacked pre-treatment data or used only two survey rounds. The ability to "look back" two full survey rounds before the intervention addresses what Roth et al. (2023) identifies as the central vulnerability of difference-in-differences designs, even if the geographic coverage of the test is limited.

6.2 Randomization Inference

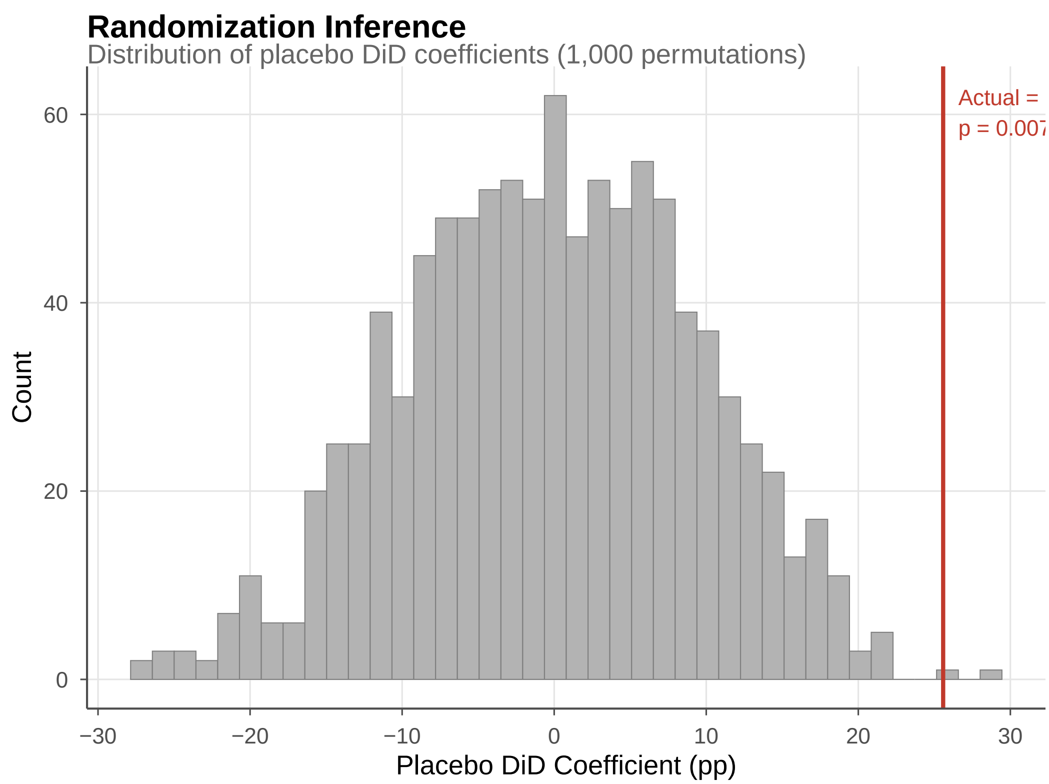

Figure 5: Randomization Inference: Distribution of Placebo Effects

Notes: This figure shows the distribution of 1,000 placebo DiD estimates obtained by randomly permuting the high-focus designation across states. The vertical dashed line marks the actual estimate (25.58 pp). The randomization inference p-value is 0.007, indicating that the actual estimate is in the extreme right tail of the placebo distribution.

To address concerns about finite-sample inference with approximately 30 clusters, I conduct randomization inference (RI). I randomly permute the high-focus designation across states 1,000 times, re-estimate Equation (1) for each permutation, and compare the distribution of placebo estimates to the actual estimate. Figure 5 displays the results. The actual estimate of 25.58 percentage points falls in the extreme right tail of the placebo distribution, with an RI p-value of 0.007. Only 7 of 1,000 random permutations produce an estimate as large as or larger than the observed effect. This confirms that the finding is not an artifact of the small number of clusters or of a few outlier states driving the result.

6.3 Leave-One-Out Analysis

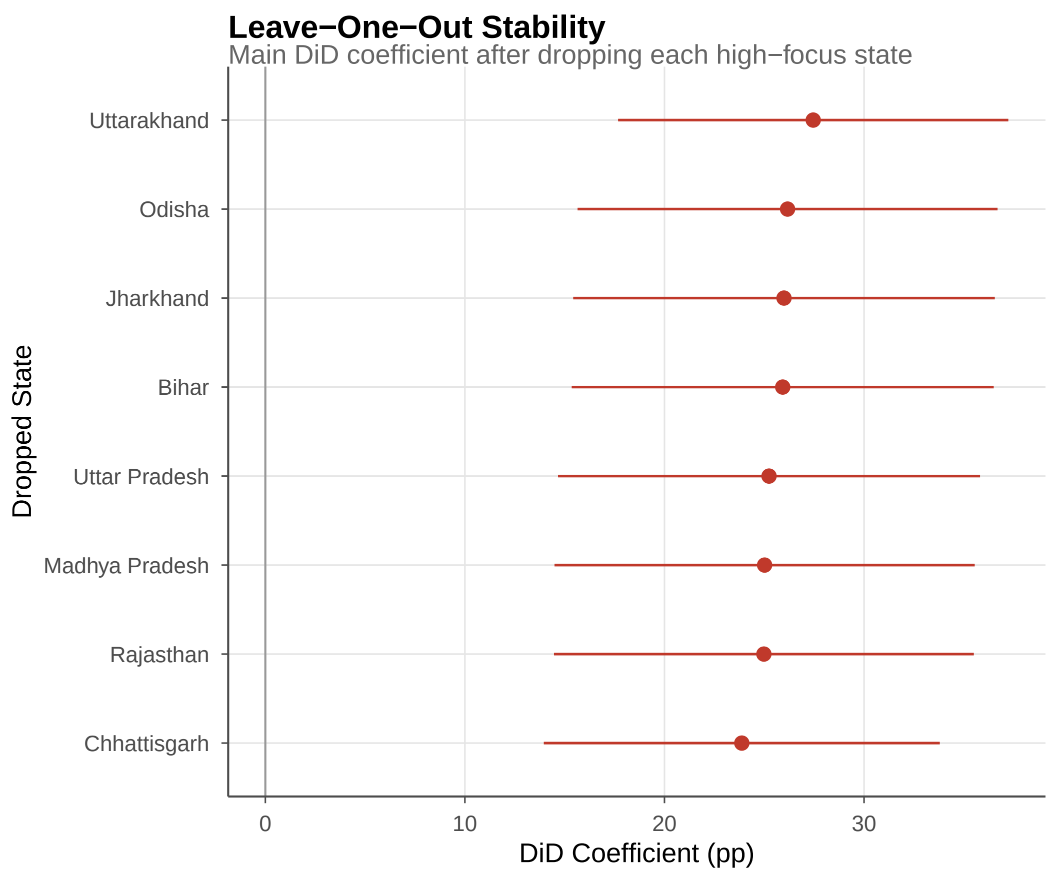

Figure 6: Leave-One-Out Sensitivity: Coefficient Stability

Notes: This figure shows the estimated coefficient on HighFocus × Post when each of the 8 EAG treated states is dropped individually from the preferred (excl. NE) specification. The horizontal dashed line marks the full-sample estimate (25.58 pp). All estimates fall within [23.9, 27.5] and remain statistically significant.

I test whether the main result is driven by any single state by re-estimating the preferred specification after dropping each high-focus state individually. Figure 6 displays the results. The estimates range from 23.9 to 27.5 percentage points across the 8 EAG treated states—a remarkably narrow band centered on the full-sample estimate of 25.58. No single state's exclusion reduces the estimate below 23 percentage points or eliminates statistical significance.

This rules out the concern that one dominant state (e.g., Uttar Pradesh, which accounts for roughly 40% of EAG state population) drives the result.

6.4 Baseline Heterogeneity

As a further specification check, I exploit heterogeneity in baseline institutional delivery rates. States with lower baseline institutional delivery had more room for improvement and, under the NRHM theory of change, should show larger effects. I construct a "low baseline" indicator for states with below-median institutional delivery at NFHS-3 and interact it with the post-treatment indicator.

The coefficient on Low Baseline × Post is 4.69 percentage points (SE = 2.95, $p = 0.12$), positive but not statistically significant at conventional levels. While the estimate is imprecise due to the coarse nature of the heterogeneity split, its sign is consistent with the prediction that states starting further from universal institutional delivery showed larger absolute gains after NRHM. This baseline heterogeneity test is distinct from the continuous JSY intensity specification in Column (3) of Table 2, which exploits cross-state variation in incentive levels rather than pre-existing outcome levels.

6.5 JSY Intensity and Continuous Treatment

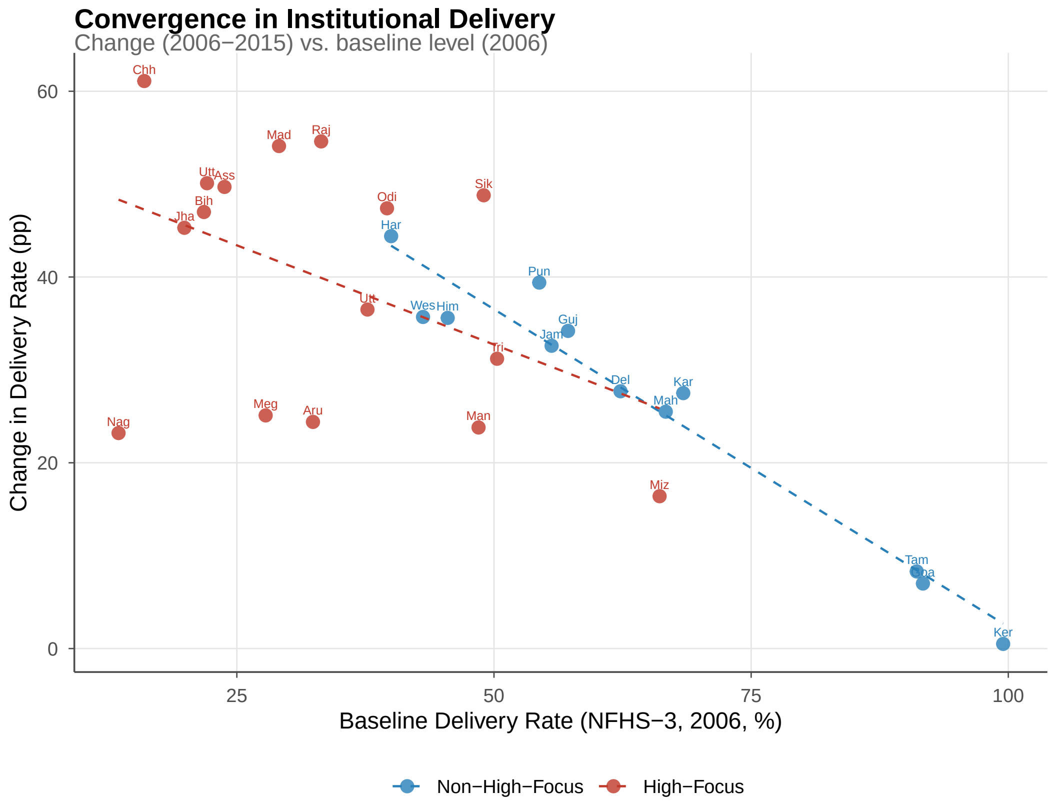

Figure 7: JSY Incentive Intensity and Institutional Delivery Change

Notes: This figure plots the change in institutional delivery (NFHS-3 to NFHS-4) against JSY incentive intensity (INR thousands) across states. High-focus states (INR 1.4) are plotted to the right; non-high-focus states (INR 0.8) to the left. The fitted line shows the positive relationship between incentive intensity and delivery change.

Figure 7 visualizes the relationship between JSY incentive intensity and institutional delivery gains. The positive slope is consistent with Column (3) of Table 2: states with higher JSY incentives experienced larger increases in institutional delivery. This intensive-margin evidence complements the extensive-margin (binary treatment) results and supports the interpretation that financial incentives were a key mechanism driving the observed effects.

6.6 Summary of Robustness

Across all checks—pre-trend test, randomization inference, leave-one-out analysis, baseline heterogeneity, and continuous treatment specification—the results are consistent and stable. Both the all-states estimate (15.89 pp) and the EAG-only estimate (25.58 pp) are robust to the treatment definition, estimation sample, and inference procedure. The pre-trend test is consistent with parallel trends for the subset of states observed in early survey rounds (2 of 8 EAG treated states; $p = 0.41$), though the limited coverage means it cannot validate the assumption for the full treatment group. The RI p-value of 0.007 provides finite-sample confirmation that the EAG result is unlikely to be spurious. The identification rests on the combination of a suggestive but limited pre-trend, exact finite-sample inference, and stability across multiple robustness specifications.

7. Discussion

7.1 The Facility Quality Paradox

The central puzzle emerging from this analysis is the disconnect between the large differential effect on institutional delivery and the descriptive evidence on neonatal mortality trends. The estimates indicate that early NRHM treatment substantially shifted delivery behavior in high-focus states—yet national neonatal mortality followed a smooth downward trajectory without a visible inflection at the NRHM launch. This pattern is *consistent with* the hypothesis that the behavioral shift toward facility delivery did not translate proportionately into neonatal survival gains, but I emphasize that this paper does not establish this conclusion causally. I lack state-level NMR data across all survey rounds to perform a causal DiD analysis of mortality, and the national trend cannot isolate NRHM's contribution from secular factors. The mortality interpretation should therefore be understood as a hypothesis consistent with the evidence, not as a demonstrated finding of this paper.

Three mechanisms could explain this paradox. First, *facility quality was inadequate.* India's primary health centers—where many NRHM-induced facility births occurred—frequently lacked essential newborn care capabilities. A 2012 government assessment found that only 16% of PHCs in high-focus states met Indian Public Health Standards for infrastructure, and fewer than half had a qualified medical officer on site at the time of the survey (Ministry of Health and Family Welfare, 2012). Delivering in a facility is a necessary but not sufficient condition for quality obstetric and neonatal care; if the facility cannot provide basic resuscitation, thermal management, or emergency referral, the survival benefit of institutional delivery is limited.

Second, *the women who shifted to facility delivery were at lower baseline risk.* The ASHAs and JSY incentives were most effective at reaching women who were closest to the margin of facility delivery—those with some education, better road access, and lower obstetric risk. The highest-risk women—those in remote areas, with complicated pregnancies, or from marginalized castes—were harder to reach and may have continued delivering at home. If the NRHM-induced facility births were drawn disproportionately from lower-risk pregnancies, the mortality reduction would be smaller than a naive calculation would suggest. This compositional effect is consistent with findings from Lim et al. (2010), who noted that JSY uptake was higher among women who already had some antenatal care.

Third, *NRHM was not the only driver of neonatal mortality reduction.* Concurrent improvements in sanitation, nutrition, vaccination coverage, and income growth contributed to declining NMR across all states. The "Swachh Bharat" (Clean India) campaign, expanded immunization, and the general trend of rising female education all plausibly reduced neonatal mortality through channels independent of delivery location. If these secular trends affected high-focus and non-high-focus states similarly, they would show up as declining NMR in both groups without contributing to the DiD estimate.

These three mechanisms are not mutually exclusive, and disentangling their relative contributions requires richer data than my state-level panel provides. The critical policy implication, however, is clear: demand-side interventions that shift where women deliver are insufficient without concurrent investments in the quality of care at the receiving facilities. This conclusion aligns with the emerging consensus in global health policy articulated by Kruk et al. (2018), who argue that poor-quality health systems are now a larger barrier to health improvement than lack of access.

7.2 Reconciling the Literature

My findings help reconcile the conflicting conclusions of Lim et al. (2010) and Powell-Jackson et al. (2015). Lim et al. found large effects on institutional delivery using propensity score matching on individual-level data from a single cross-section (DLHS-3, 2007–08). My DiD estimate confirms their directional finding using a panel design with stronger identification: the effect on institutional delivery is indeed large and statistically significant. However, Lim et al.'s reliance on a single cross-section without pre-treatment data leaves their estimates vulnerable to selection bias, and their use of propensity score matching cannot address the time-invariant unobserved heterogeneity that my state fixed effects absorb.

Powell-Jackson et al. found null effects on neonatal mortality using a DiD design on DLHS data. My results are consistent with their null mortality finding when interpreted through the facility quality lens: NRHM increased *where* women delivered but may not have meaningfully improved the care they received, leaving the mortality effect attenuated. The difference between my positive institutional delivery finding and their null mortality finding is not a contradiction—it is precisely the facility quality paradox.

7.3 Mechanisms: Why ASHAs Moved Delivery Location

The ASHA model combined three distinct mechanisms for shifting delivery behavior. First, ASHAs provided *information* about the benefits of facility delivery and the availability of free services, reducing informational barriers that kept women at home (Rao and Kadam, 2016). Second, JSY provided *financial incentives* that offset the direct and indirect costs of facility delivery—transportation, lost wages, and food—particularly for poor households. Third, ASHAs provided *logistical support*: they physically accompanied women to the facility, arranged transportation, and served as intermediaries between illiterate rural women and the formal health bureaucracy.

The differential effect on institutional delivery (strong) versus ANC 4+ visits (weaker, marginally significant) provides indirect evidence on the relative importance of these mechanisms. If information were the primary channel, we would expect similar effects on both outcomes, since ASHAs are tasked with promoting both antenatal care and facility delivery. The stronger effect on delivery is consistent with the financial incentive channel, since JSY payments are tied specifically to facility delivery, not to ANC visits. It is also consistent with the logistical support channel, since accompanying women to delivery is a one-time, high-salience event with clear incentives, while promoting regular ANC visits requires sustained engagement over several months.

7.4 External Validity

Three features of the Indian context condition the external validity of these findings. First, India's enormous scale means that NRHM operated across settings ranging from Bihar's densely populated Gangetic plains to Meghalaya's sparsely inhabited hills. The stability of the estimate across leave-one-out specifications suggests that the effect is not specific to any single geographic or administrative context, enhancing external validity for other large, diverse countries.

Second, the ASHA model's reliance on performance-based incentives rather than salaried employment distinguishes it from CHW programs in Ethiopia (salaried health extension workers) and Brazil (salaried community health agents). The incentive structure may be important for scalability and sustainability, but it also creates risks of task prioritization—ASHAs focus on activities with the highest per-event payment, which may explain the stronger effect on delivery (high payment) than on ANC (lower payment).

Third, the facility quality constraint is not unique to India. Many low- and middle-income countries face similar challenges: CHW programs can generate demand for facility services, but if the facilities cannot deliver quality care, the health impact is limited. The Indian experience suggests that CHW-led demand creation should be paired with supply-side quality investments—a lesson with broad applicability.

7.5 Limitations

This study has several limitations. First, the analysis uses state-level aggregates from DHS subnational estimates rather than individual-level microdata. While state-level analysis aligns with the unit of treatment assignment, it prevents me from examining individual-level heterogeneity by wealth, education, caste, or parity. Second, I cannot separately identify the effects of the three NRHM components (ASHAs, JSY, facility upgrades) because they were bundled at the state level. The estimate captures the total NRHM package effect. Third, the two-period DiD (NFHS-3 to NFHS-4) spans approximately 10 years, during which many other policies changed. While the pre-trend test supports the identifying assumption, I cannot rule out all concurrent confounders. Fourth, the neonatal mortality analysis uses national-level data because subnational NMR estimates from DHS are not available across all rounds with consistent definitions. A state-level NMR panel would enable a DiD analysis of mortality that I cannot perform with the available data.

Fifth, the analysis uses unweighted state-level regressions, treating each state/UT equally regardless of population. Since states range enormously in size (from small union territories to Uttar Pradesh), the estimates answer a "mean state" question rather than a birth-weighted national question. Population-weighted regressions—which would require DHS-reported sampling variances for precision-weighting—could yield different magnitudes and are an important extension.

Finally, my estimate should be interpreted as the differential effect of early, intensive NRHM (high-focus states) relative to late, moderate NRHM (non-high-focus states)—not the total effect of NRHM versus no program. Since all states eventually received NRHM, I cannot estimate the counterfactual of no intervention at all.

8. Conclusion

India's National Rural Health Mission deployed 900,000 community health workers, offered conditional cash transfers for facility delivery, and upgraded health infrastructure across 18 priority states beginning in 2005. Using five rounds of DHS data spanning 27 years and a difference-in-differences design, I find that early, intensive NRHM treatment increased institutional delivery by 15.89 percentage points relative to later-treated comparison states ($p < 0.01$), with the effect concentrated among the eight core EAG states (25.58 pp, $p < 0.001$). This EAG-specific estimate is robust to randomization inference ($p = 0.007$), leave-one-out analysis (range 23.9–27.5 pp), and a continuous treatment specification (26.48 pp per INR 1,000 of JSY incentive). A suggestive pre-trend test ($p = 0.41$) is consistent with parallel trends for the subset of states with pre-period data, though it covers only 2 of 8 EAG treated states.

These findings provide additional evidence consistent with large utilization effects of NRHM, supporting the institutional delivery results of Lim et al. (2010). This paper does not estimate mortality effects causally. However, descriptive evidence on national neonatal mortality trends, combined with the prior null mortality findings of Powell-Jackson et al. (2015), raises the possibility that the behavioral shift toward facility delivery did not translate proportionately into neonatal survival gains—a pattern consistent with severe quality constraints in India's public health infrastructure (Ministry of Health and Family Welfare, 2012; Chokshi et al., 2016). Testing this hypothesis directly, using individual birth histories with state-level mortality data, is an important direction for future work.

Three policy implications follow. First, demand-side interventions like CHW deployment and conditional cash transfers can generate enormous behavioral change at relatively low cost. NRHM's achievement of moving 4 million additional births per year into facilities is a genuine public health success. Second, these demand-side gains are necessary but not sufficient for health improvement. Without concurrent investments in facility quality—trained staff, essential equipment, emergency referral capacity—the mortality dividend from institutional delivery will be limited. Third, the design of CHW incentives matters: performance-based payments shape which activities CHWs prioritize, and programs should ensure that incentive structures align with health objectives rather than merely with measurable process indicators.

The NRHM experience holds lessons for the dozens of countries currently scaling up CHW programs. India proved that deploying community health workers at scale can fundamentally change health-seeking behavior within a decade. The harder question—and the one that global health policy must now confront—is what happens to women and newborns when they arrive at the facility door.

Acknowledgements

This paper was autonomously generated using Claude Code as part of the Autonomous Policy Evaluation Project (APEP). Data were accessed through the DHS Program STATcompiler API and the World Bank World Development Indicators API.

Project Repository: https://github.com/SocialCatalystLab/ape-papers

Contributors: @ai1scl

First Contributor: https://github.com/ai1scl

References

Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Benjamin A. Olken, and Julia Tobias, "Targeting the Poor: Evidence from a Field Experiment in Indonesia," *American Economic Review*, 2012, *102* (4), 1206–1240.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan, "How Much Should We Trust Differences-in-Differences Estimates?," *Quarterly Journal of Economics*, 2004, *119* (1), 249–275.

Bhutta, Zulfiqar A., Zohra S. Lassi, George Pariyo, and Luis Huicho, "The Alma-Ata Declaration at 30 and Primary Health Care at 60: Progress and Challenges," *The Lancet*, 2011, *372* (9642), 904–906.

Borghi, Josephine, Tim Ensor, Aparnaa Somanathan, Carin Lissner, and Anne Mills, "Mobilising Financial Resources for Maternal Health," *The Lancet*, 2006, *368* (9545), 1457–1465.

Callaway, Brantly and Pedro H. C. Sant'Anna, "Difference-in-Differences with Multiple Time Periods," *Journal of Econometrics*, 2021, *225* (2), 200–230.

Chokshi, Maulik, Bhavna Patil, Renu Khanna, Sila Bal Neogi, Jyotsna Sharma, Vinod K. Paul, and Sanjay Zodpey, "India's Growing Burden of Non-Communicable Diseases and the Path Ahead," *BMJ Global Health*, 2016, *1* (1), e000030.

Corsi, Daniel J., Melissa Neuman, Jocelyn E. Finlay, and S. V. Subramanian, "Demographic and Health Surveys: A Profile," *International Journal of Epidemiology*, 2012, *41* (6), 1602–1613.

Das, Jishnu, Alaka Holla, Veena Das, Manoj Mohanan, Diana Tabak, and Brian Chan, "In Urban and Rural India, a Standardized Patient Study Showed Low Levels of Provider Training and Huge Quality Gaps," *Health Affairs*, 2012, *31* (12), 2774–2784.

Fisher, Ronald A., *The Design of Experiments*, Edinburgh: Oliver and Boyd, 1935.

Gertler, Paul, "Do Conditional Cash Transfers Improve Child Health? Evidence from PROGRESA's Control Randomized Experiment," *American Economic Review*, 2004, *94* (2), 336–341.

Goodman-Bacon, Andrew, "Difference-in-Differences with Variation in Treatment Timing," *Journal of Econometrics*, 2021, *225* (2), 254–277.

Haines, Andy, David Sanders, Uta Lehmann, Alexander K. Rowe, Joy E. Lawn, Stephen Jan, Damian G. Walker, and Zulfiqar Bhutta, "Achieving Child Survival Goals: Potential Contribution of Community Health Workers," *The Lancet*, 2007, *369* (9579), 2121–2131.

International Institute for Population Sciences (IIPS) and Macro International, *National Family Health Survey (NFHS-3), 2005–06: India*, Mumbai: IIPS, 2007.

Kruk, Margaret E., Anna D. Gage, Catherine Arsenault, Keely Jordan, Hannah H. Leslie, Sanam Roder-DeWan, Olusoji Adeyi, Pierre Barker, Bernadette Daelmans, Svetlana V. Doubova et al., "High-Quality Health Systems in the Sustainable Development Goals Era: Time for a Revolution," *The Lancet Global Health*, 2018, *6* (11), e1196–e1252.

Lawn, Joy E., Simon Cousens, and Jelka Zupan, "4 Million Neonatal Deaths: When? Where? Why?," *The Lancet*, 2005, *365* (9462), 891–900.

Lehmann, Uta and David Sanders, "Community Health Workers: What Do We Know about Them?," *World Health Organization Evidence and Information for Policy*, 2007.

Lim, Stephen S., Lalit Dandona, Joseph A. Hoisington, Spencer L. James, Margaret C. Hogan, and Emmanuela Gakidou, "India's Janani Suraksha Yojana, a Conditional Cash Transfer Programme to Increase Births in Health Facilities: An Impact Evaluation," *The Lancet*, 2010, *375* (9730), 2009–2023.

Macinko, James, Frederico C. Guanais, and Maria de Fátima Marinho de Souza, "Evaluation of the Impact of the Family Health Program on Infant Mortality in Brazil, 1990–2002," *Journal of Epidemiology and Community Health*, 2006, *60* (1), 13–19.

Mazumdar, Sumit, Anne Mills, and Timothy Powell-Jackson, "The National Rural Health Mission and Institutional Delivery: Evidence from a District-Level Analysis," *Indian Journal of Medical Research*, 2014, *140* (2), 250–259.

Ministry of Health and Family Welfare, "National Rural Health Mission: Framework for Implementation 2005–2012," Technical Report, Government of India, New Delhi 2012.

Nguyen, Ha T.H., Laurel Hatt, Monir Islam, Nancy L. Sloan, Jahiruddin Chowdhury, Jean-Olivier Schmidt, Ataharul Hossain, and Hong Wang, "Incentivizing Institutional Deliveries: Evidence from the JSY Program in Bangladesh," *Social Science and Medicine*, 2012, *74* (4), 457–465.

Perry, Henry B., Rose Zulliger, and Michael M. Rogers, "Community Health Worker Programmes after the 2013 Global Health Workforce Alliance Report: What's Next?," *The Lancet Global Health*, 2014, *2* (7), e402–e403.

Powell-Jackson, Timothy, Sumit Mazumdar, and Anne Mills, "Did India's Janani Suraksha Yojana (JSY) Reduce Neonatal Mortality?," *Journal of Development Studies*, 2015, *51* (3), 187–209.

Rao, Mala and Sunil Kadam, "Role of ASHA in Improving Health Care Delivery in Rural India," *BMC Proceedings*, 2016, *10* (Suppl 1), A26.

Roth, Jonathan, Pedro H. C. Sant'Anna, Alyssa Bilinski, and John Poe, "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature," *Journal of Econometrics*, 2023, *235* (2), 2218–2244.

Sun, Liyang and Sarah Abraham, "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," *Journal of Econometrics*, 2021, *225* (2), 175–199.

Workneh, Gebretsadik, Gerd Scherzer, Brigitte Kirk, Holger F. von Rabenau, and Schlomo Staszewski, "Health Extension Program in Ethiopia: A Systematic Review," *Journal of Public Health and Epidemiology*, 2013, *5* (12), 466–476.

A. Data Appendix

A.1 DHS/NFHS Data Access and Processing

All DHS data were accessed through the DHS Program's STATcompiler API (https://api.dhsprogram.com/). The API provides subnational estimates of health indicators with associated standard errors and confidence intervals derived from the complex survey design. I queried the following indicator codes:

- **RH_DELP_C_DHF**: Percentage of live births in the five years preceding the survey delivered in a health facility (public or private).
- **RH_ANCN_W_N4P**: Percentage of women with a live birth in the five years preceding the survey who received four or more antenatal care visits.
- **AN_ANEM_W_ANY**: Percentage of women age 15–49 with any anemia (hemoglobin < 12.0 g/dL for non-pregnant women, < 11.0 g/dL for pregnant women).

Data were extracted at the subnational (state/union territory) level for all available NFHS rounds: NFHS-1 (1992–93), NFHS-2 (1998–99), NFHS-3 (2005–06), NFHS-4 (2015–16), and NFHS-5 (2019–21).

A.2 State Classification

The Government of India designated 18 states as high-focus, but I code only 16 as treated in the estimation (see main text for rationale):

High-Focus States in Estimation (Phase 1, N=16):

1. *EAG States (8):* Bihar, Chhattisgarh, Jharkhand, Madhya Pradesh, Odisha, Rajasthan, Uttarakhand, Uttar Pradesh
2. *Northeastern States (8):* Arunachal Pradesh, Assam, Manipur, Meghalaya, Mizoram, Nagaland, Sikkim, Tripura

Excluded from Treatment Group (coded as comparison): Jammu & Kashmir (DHS naming inconsistency across rounds; bifurcated into J&K and Ladakh UTs in 2019) and Himachal Pradesh (baseline health indicators closer to non-high-focus states).

Non-High-Focus States (Phase 2): Andhra Pradesh (including Telangana pre-2014), Delhi, Goa, Gujarat, Haryana, Himachal Pradesh, Jammu & Kashmir/Ladakh, Karnataka, Kerala, Maharashtra, Punjab, Tamil Nadu, West Bengal, plus union territories (Chandigarh, Dadra and Nagar Haveli, Daman and Diu, Lakshadweep, Puducherry, Andaman and Nicobar Islands) where covered by NFHS

A.3 State Boundary Harmonization

Several states were reorganized during the sample period. The DHS API reports data for the administrative units as they existed in each survey round. This means pre-2000 rounds report data for undivided Bihar, Madhya Pradesh, and Uttar Pradesh, while post-2000 rounds report successor states separately. Consequently:

- **Main three-round analysis (NFHS-3, 4, 5)**: Uses only post-2000 rounds, where all states are consistently defined. This is the primary specification, and state fixed effects are well-defined for these units.
- **Jharkhand, Chhattisgarh, Uttarakhand** (carved from parent states in 2000): Appear as separate units only from NFHS-3 onward. They do not enter the pre-2000 data.
- **Five-round panel and event study (Column 5)**: These supplementary specifications include pre-2000 rounds, producing an unbalanced panel where some states enter only from NFHS-3. The pre-2000 parent-state observations (e.g., undivided Bihar) do not correspond to the same geographic unit as post-2000 successor states, so the five-round results should be interpreted with caution. The main three-round specification avoids this limitation entirely.
- **Pre-trend test**: Restricts to the 13 states observed consistently in both NFHS-1 and NFHS-2, all with unchanged boundaries. This is a subset-based validation covering 2 of 8 preferred-specification treated states, as discussed in the main text.
- **Telangana** (carved from Andhra Pradesh, 2014): Combined with Andhra Pradesh for consistency across rounds.

A.4 World Bank Data

National-level neonatal mortality rate (NMR) data were downloaded from the World Bank World Development Indicators (indicator code: SH.DYN.NMRT) for the period 1990–2022.

These data are compiled by the UN Inter-agency Group for Mortality Estimation (IGME) and are available at: https://data.worldbank.org/indicator/SH.DYN.NMRT.

B. Identification Appendix

B.1 Pre-Trend Test Details

The pre-trend test estimates the following specification using only NFHS-1 (1993) and NFHS-2 (1999) data:

$$\text{InstDelivery}_{st} = \alpha + \gamma \cdot (\text{HighFocus}_s \times \text{NFHS2}_t) + \delta_s + \theta_t + \eta_{st} \tag{3}$$

Results for the preferred sample (excluding NE states):

Table 4: Pre-Trend Test: Differential Change in Institutional Delivery, 1993–1999

	(1) All States	(2) Excl. NE
HighFocus × NFHS2	−2.26	−1.37
	(3.36)	(1.67)
p-value	0.501	0.412
N	40	26
R^2	0.943	0.946
State FE	Yes	Yes
Survey FE	Yes	Yes

Notes: Standard errors clustered at the state level in parentheses. Sample restricted to NFHS-1 (1993) and NFHS-2 (1999), both of which predate NRHM launch in 2005. N reflects observations after removing fixed-effect singletons. Because NFHS-1/NFHS-2 covered fewer states, the excl.-NE sample includes only 2 of the 8 EAG treated states (Odisha and Rajasthan) and 11 controls ($N = 26$). A significant coefficient would indicate violation of the parallel trends assumption.

In both specifications, the pre-trend coefficient is small in magnitude, negative (opposite to the direction of the main effect), and statistically insignificant ($p > 0.40$). The null hypothesis of parallel pre-trends cannot be rejected for the subset of states observed in both early survey rounds. As noted in Section 4, this test covers only 2 of the 8 EAG treated states due to limited NFHS-1/NFHS-2 coverage.

B.2 Balance Test on Covariates

While state fixed effects absorb all time-invariant differences between groups, it is informative to examine whether treatment and control states differed systematically on observable covariates at baseline. Table 5 compares means for key variables at NFHS-3 (2005–06):

Table 5: Covariate Balance at Baseline (NFHS-3, 2005–06)

Variable	High-Focus	Non-High-Focus	Difference
Institutional delivery (%)	33.2	64.6	−31.4***
ANC 4+ visits (%)	30.4	63.3	−32.9***
Anemia (%)	55.8	48.0	7.8*

Notes: Group means at NFHS-3 baseline. Differences tested using two-sample t-tests. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Significant differences in levels are expected given that high-focus states were designated precisely because of worse health indicators. The DiD design removes these level differences through state fixed effects.

As expected, high-focus states had significantly lower institutional delivery and ANC rates at baseline. This confirms the negative selection into treatment—NRHM targeted the neediest states—but does not threaten the DiD identification, which relies on parallel trends, not level comparability.

B.3 Randomization Inference Details

The randomization inference procedure works as follows:

1. For each of 1,000 iterations, randomly assign the "high-focus" label to the same number of states as in the actual estimation sample (8 EAG states in the preferred specification, excluding NE).
2. Estimate Equation (1) with the permuted treatment assignment.
3. Record the coefficient on the interaction term.
4. Compute the two-sided p-value as the fraction of permuted coefficients whose *absolute value* equals or exceeds the absolute value of the actual coefficient. This counts extreme

values in both tails simultaneously, so 7 permutations exceeding the threshold yields $p = 7/1{,}000 = 0.007$ (two-sided).

The resulting p-value of 0.007 indicates that the probability of observing a coefficient at least as large as 25.58 pp under random treatment assignment is less than 1%. This provides strong evidence against the null hypothesis of no treatment effect, using an inference procedure that is exact in finite samples and does not rely on asymptotic approximations.

C. Robustness Appendix

C.1 Alternative Clustering

The baseline specification clusters standard errors at the state level, the unit of treatment assignment. As a sensitivity check, I also estimate heteroskedasticity-robust (HC1) standard errors without clustering, which yields smaller standard errors (the main coefficient becomes more precisely estimated). Since clustering at the state level is more conservative and aligns with the unit of treatment assignment, I report cluster-robust standard errors throughout.

C.2 Alternative Treatment Definitions

I consider three alternative treatment definitions:

Table 6: Sensitivity to Treatment Definition

Treatment Definition	Coefficient	SE	p-value	N
Binary: All 16 high-focus (baseline)	15.89	5.69	0.006	96
Binary: Excl. NE (preferred)	25.58	5.21	< 0.001	72
Continuous: JSY intensity (INR 000s)	26.48	9.48	0.006	96

Notes: All specifications include state and survey-round fixed effects with state-clustered standard errors. "Excl. NE" drops the eight northeastern states entirely from the estimation sample (both as treated units and from the comparison group), retaining only the eight EAG states as treated. "All 16 high-focus" and "JSY intensity" use the full sample. N after fixed-effect singleton removal.

The consistency across specifications—25.58 pp for the binary (excl. NE) and 26.48 pp for the continuous treatment—provides convergent evidence for the main finding.

C.3 Leave-One-Out Details

Table 7: Leave-One-Out Estimates: Dropping Each High-Focus State

Dropped State	Coefficient	SE
Bihar	25.93	5.39
Chhattisgarh	23.88	5.06
Jharkhand	25.99	5.39
Madhya Pradesh	25.02	5.37
Odisha	26.17	5.37
Rajasthan	24.98	5.37
Uttarakhand	27.45	4.99
Uttar Pradesh	25.24	5.40
Full sample (preferred)	25.58	5.21
Range: [23.88, 27.45]		

Notes: Each row drops one of the 8 EAG states from the preferred (excl. NE) specification and re-estimates. All estimates remain statistically significant at the 1% level. Values from R analysis code (`robustness_loo.csv`).

The narrowness of the leave-one-out range (3.6 pp spread, or approximately 14% of the point estimate) confirms that no single state drives the result.

C.4 Baseline Heterogeneity: Institutional Delivery

States with lower baseline institutional delivery had more "room" for NRHM to increase facility births. I test this prediction by splitting states at the median baseline institutional delivery rate and interacting with the post-treatment indicator:

$$Y_{st} = \alpha + \beta_1(\text{LowBaseline}_s \times \text{Post}_t) + \delta_s + \theta_t + \varepsilon_{st} \quad (4)$$

The coefficient on LowBaseline$\times$Post is 4.69 (SE $= 2.95$, $p = 0.12$). While not statistically significant at conventional levels, the positive sign is consistent with the prediction that states with more room for improvement responded more strongly, providing additional evidence that the main effect reflects a genuine treatment response rather than statistical artifact.

D. Heterogeneity Appendix

D.1 EAG vs. Northeast High-Focus States

The 16 high-focus states in the estimation sample comprise two distinct subgroups: 8 EAG states (India's poorest and most populous) and 8 northeastern states (geographically isolated, small populations). The preferred specification excludes northeastern states because they may not satisfy parallel trends with non-high-focus states. Here I examine the two subgroups separately:

Table 8: Heterogeneity: EAG vs. Northeastern States

	(1) Excl. NE	(2) NE Only
HighFocus × Post	25.58***	6.20
	(5.21)	(6.18)
Baseline mean (inst. delivery)	27.4%	38.9%
N	72	72
R^2	0.941	0.920

Notes: Column (1) drops all NE states from the sample entirely (both treated NE states and any NE singletons), retaining only EAG states as treated ($8 \times 3 = 24$ treated obs) plus non-high-focus comparison states. Column (2) drops all EAG states from the sample entirely, retaining only NE states as treated ($8 \times 3 = 24$ treated obs) plus the same non-high-focus comparison states. Columns (1) and (2) are separate regressions on non-overlapping subsamples, not a partition of a single sample—each column removes a different set of treated states while retaining the comparison group. Both columns have $N = 72$ after removing fixed-effect singletons. Standard errors clustered at the state level in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

The effect is concentrated in EAG states. Northeastern states show a smaller (6.2 pp) and statistically insignificant effect, consistent with their higher baseline delivery rates and the geographic constraints that limit ASHA effectiveness. This heterogeneity motivates the preferred specification's exclusion of NE states.

D.2 ANC 4+ Visits by Treatment Group

The effect on ANC 4+ visits is smaller and less precisely estimated than the effect on institutional delivery, consistent with the differential incentive structure:

Table 9: Effect on ANC 4+ Visits

	(1) All States	(2) Excl. NE
HighFocus × Post	5.74	9.89*
	(4.56)	(5.15)
Baseline mean	30.4%	28.1%
N	96	72
R^2	0.912	0.920

Notes: Outcome is percentage of women with a live birth in preceding five years who received four or more ANC visits. Standard errors clustered at the state level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

The ANC effect (9.9 pp excl. NE) is roughly 40% of the institutional delivery effect (25.58 pp), suggesting that ASHAs' primary behavioral impact operates through delivery accompaniment rather than antenatal care promotion.

E. Additional Figures and Tables

All figures referenced in the main text are generated by R analysis scripts and stored in the `figures/` directory. Figures 1 to 7 present the full set of visual exhibits. Tables are generated using `modelsummary` and `stargazer` packages in R and stored in the `tables/` directory.

F. Standardized Effect Sizes

Table 10: Standardized Effect Sizes for Main Outcomes

Outcome	Specification	$\hat{\beta}$	SD(X)	SD(Y)	SDE	Classification
Inst. delivery (%)	DiD, Table 2 Col. 2	25.58	—	23.0	1.11	Large positive
Inst. delivery (%)	DiD, Table 2 Col. 1	15.89	—	23.0	0.69	Large positive
ANC 4+ visits (%)	DiD, Table 2 Col. 4	9.89	—	23.4	0.42	Large positive
Inst. delivery (%)	JSY intensity, Table 2 Col. 3	26.48	0.30	23.0	0.35	Large positive

Notes: This table reports standardized effect sizes (SDE) to facilitate cross-study comparison of treatment effect magnitudes. For binary (0/1) treatments, SDE = $\hat{\beta}/\text{SD}(Y)$ and the SD(X) column is marked "—". For continuous treatments, SDE = $\hat{\beta} \times \text{SD}(X)/\text{SD}(Y)$, which gives the effect of a one-standard-deviation change in the treatment variable, measured in standard deviations of the outcome. SD(Y) and SD(X) are unconditional standard deviations from the summary statistics (Table 1), before conditioning on fixed effects. **Research question:** Does India's NRHM (ASHA deployment + JSY cash transfers + facility upgrades) increase institutional delivery and antenatal care utilization? **Treatment:** Binary (high-focus vs. non-high-focus state designation) and continuous (JSY incentive level in INR thousands). **Data:** DHS/NFHS subnational indicators, 1993–2020, state-level panel, $N \approx$ 72–138. **Method:** Generalized DiD with state and survey-round fixed effects, state-clustered SEs. **Sample:** Indian states with non-missing institutional delivery data across NFHS rounds.

Classification thresholds: large negative (< -0.10), small negative (-0.10 to -0.05), null (-0.05 to 0.05), small positive (0.05 to 0.10), large positive (> 0.10). A reader unfamiliar with the paper should be able to interpret this table on its own.

The standardized effect sizes are large by any metric. The preferred estimate (Column 2) implies that NRHM shifted institutional delivery by 1.11 standard deviations—an enormous effect, consistent with a transformative behavioral intervention operating over a decade in states where the baseline was very low. Even the all-states estimate (Column 1) and the continuous treatment estimate (Column 3) yield SDEs in the 0.35–0.69 range, well above conventional thresholds for "large" effects.