



District-Level Healthcare Infrastructure and Outcome Gaps in India: An Empirical Analysis of Secondary Care Specialist Shortfalls and Primary Infrastructure

¹Kaustubh R. Lipare, ²Najuka Jadhav, ³Kaavya Parashar, ⁴Enugandula Gangasairam, ⁵Indu Bharti, ⁶Daksh Pokar

¹⁻⁶International Institute of SDGs and Public Policy Research (IISPPR)

ABSTRACT

Substantial public financial investments under the National Health Mission (NHM) have expanded physical health infrastructure across India. However, distinct district-level disparities in maternal and child health outcomes persist. This study evaluates whether variations in the Infant Mortality Rate (IMR) are driven primarily by deficits in physical infrastructure ('bricks') or critical structural shortages of clinical specialists ('brains') at referral centers. Utilizing a quantitative, cross-sectional, and ecological design, we analyzed a stratified systematic sample of 80 administrative districts from two Empowered Action Group (EAG) states (Uttar Pradesh and Chhattisgarh) and two Non-EAG states (Maharashtra and Tamil Nadu). Multivariate Ordinary Least Squares (OLS) regression models evaluated the predictive impact of primary clinic density, public hospital bed density, specialist doctor vacancy rates, and private sector delivery shares on district IMR, while controlling for female literacy and multidimensional poverty. Physical primary clinic density is relatively uniform across study regions and approaches standard Indian Public Health Standards (IPHS). Conversely, acute specialist physician vacancies at Community Health Centres (CHCs) average 79.1% in EAG districts compared to 36.5% in Non-EAG cohorts. Multivariate regression models reveal that the specialist doctor vacancy rate is the single most powerful statistical predictor of elevated district IMR ($p < 0.001$), carrying 5.2 times greater relative explanatory weight than primary clinic footprint ($|\beta^*| = 0.692$ vs. $|\beta^*| = 0.133$). Furthermore, a distinct moderation tipping point is observed: the marginal statistical return of constructing additional primary facilities flattens to zero once local specialist vacancies cross the 75% to 85% threshold. Diminishing returns exist for capital expenditure dedicated to building additional physical health structures if they remain unstaffed clinical shells. Policy instruments must prioritize human resource retention packages, functional secondary care, and verified institutional service readiness to accelerate progress toward regional Sustainable Development Goal (SDG) health targets.

Key words - Infant Mortality Rate; Healthcare Infrastructure; Specialist Vacancies; Indian Public Health Standards; Ecological Study; Public Policy

1. INTRODUCTION

1.1 Background and Problem: Spatial Heterogeneity in Health Outcomes

While India has achieved significant macroeconomic progress, state-level maternal and child health averages conceal a sharper reality on the ground: the physical presence of a health facility on a government register does not mean that facility is functionally equipped to



handle an obstetric or neonatal emergency. Historically, public health policy in India has prioritized the expansion of the built environment—focusing heavily on expanding the sheer volume of Sub-Centres, Primary Health Centres (PHCs), and hospital beds ([Ministry of Health and Family Welfare \[MoHFW\], 2022](#)). This persistent capital focus has treated district-level shortfalls primarily as a physical resource supply problem.

What remains far less settled is whether the more consequential bottleneck lies in the institutional readiness and clinical capacity of these existing structures. Specifically, it remains unclear whether peripheral and secondary facilities are staffed by the medical specialists required to manage complicated deliveries, and whether a functioning referral pathway exists to transfer a patient safely from a primary outpost to a center equipped to provide emergency neonatal and obstetric care. This operational tension between physical asset expansion ('bricks') and clinical human resource availability ('brains') forms the core operational puzzle of contemporary Indian healthcare delivery ([Patel & Ladusingh, 2015](#)). Therefore, a critical empirical question arises: do localized bottlenecks in clinical specialist availability and secondary-tier referral capacity function as the primary drivers of district-level infant mortality, rather than an absolute deficit in primary outpatient structures?

1.2 Magnitude of District-Level Disparity

The scale of this underlying disparity is considerable and spatially clustered across India's administrative geography ([Goli et al., 2022](#)). Approximately 448 of India's 640 districts report a maternal mortality ratio above the threshold set by the Sustainable Development Goals (SDG-3), with the most vulnerable zones concentrated across the central, eastern, and north-eastern regions ([Goli et al., 2022](#)). Infant mortality follows a similarly uneven spatial pattern, with high-mortality hotspots remaining deeply entrenched across central India and northern Uttar Pradesh even after adjusting for socioeconomic confounders and spatial autocorrelation ([Puranik et al., 2018](#); [Ranjan & Dwivedi, 2022](#)).

Within the 80-district sample analyzed in this study, the Infant Mortality Rate (IMR) ranges from 9.20 to 68.10 deaths per 1,000 live births—a seven-fold spread—while specialist doctor vacancy at Community Health Centres (CHCs) spans from 15.20% to 89.40%. Crucially, districts in Empowered Action Group (EAG) states report an average IMR that is two and a half times higher than that of non-EAG states (48.60 vs. 19.48 deaths per 1,000 live births), alongside a specialist vacancy rate that is more than double (79.06% vs. 36.51%).

This comparative cohort framework is methodologically and substantively significant: it evaluates two functionally distinct health systems operating under the exact same national policy guidelines. Investigating this regional divide allows us to determine whether these systemic outcome gaps are primarily a function of physical infrastructure layout or if they are driven by severe, localized specialist shortfalls and secondary referral voids.

1.3 Rationale and Policy Relevance

The stakes of resolving this structural puzzle are high because national capital expenditure tracks whichever explanation policymakers accept. Flagship national programmes, most notably the National Health Mission (NHM) and the Pradhan Mantri Ayushman Bharat Health Infrastructure Mission (PM-ABHIM), direct substantial public funding toward



physical facility construction on the implicit premise that expanding the built environment is the primary lever for closing regional outcome gaps ([MoHFW, 2022](#)).

If, however, the binding constraint in India's worst-performing districts is not the number of physical buildings but rather the availability of specialized clinical staffing and secondary referral capacity, continued capital investment in construction will fail to move mortality figures in high-burden districts, leading to a severe misallocation of scarce public resources. This study directly tests the hypothesis that physical primary structures yield diminishing returns on infant survival when they are decoupled from functional secondary-care specialists, providing a necessary empirical baseline to guide national health infrastructure allocation.

1.4 The Research Gap

The existing literature fails to resolve this tension due to fragmented study designs. Public health research tends to examine either socioeconomic determinants—such as rural-urban coverage gaps and maternal literacy ([Panda et al., 2020](#))—or facility-level physical infrastructure availability in isolation, rarely modeling their intersection. Where the built environment is evaluated, it is typically treated as a binary presence (i.e., whether a facility exists or not), rather than through a graded metric of functional service readiness ([Sharma et al., 2018](#)).

The primary novelty of this study lies in its integrated econometric modeling. By entering primary clinic density, public hospital bed density, and secondary specialist doctor vacancy rates into a single multivariate regression model, this study isolates physical infrastructure from clinical human capital. Furthermore, by introducing an explicit statistical interaction term between clinic density and specialist vacancy, this design tests a critical gap left open by prior work: whether the protective association of primary facility expansion collapses entirely once specialist staffing shortages pass a critical operational threshold. Consequently, the exact objective arising from this gap is to evaluate the interactive and independent associations of physical infrastructure density and specialist workforce shortages with district-level infant mortality across India.

1.5 Aim and Research Questions

This study adopts a quantitative, cross-sectional, comparative ecological design to evaluate supply-side healthcare delivery bottlenecks across diverse institutional cohorts in India. The investigation is guided by three formal research questions:

RQ1: To what extent do existing district-level physical clinic structures and secondary clinical specialist cadres deviate from statutory Indian Public Health Standards (IPHS) benchmarks (evaluating population-based norms for physical facilities and sanctioned per-facility staffing norms for specialist physicians)?

RQ2: Does primary physical clinic density or secondary-tier specialist doctor availability exert a stronger statistical association with district-level infant mortality rates?



RQ3: Does the explanatory association of clinical specialist shortages with infant mortality persist after controlling for district-level multidimensional poverty, female literacy, private-sector delivery shares, and structural regional cohort effects?

1.6 Study Objectives

To systematically answer these research questions, the study executes the following action-oriented objectives:

- **Quantify Gaps Against Statutory Standards:** Measure the absolute and percentage deviations of primary clinics, public inpatient beds, and specialist postings against official IPHS benchmarks (evaluating population-based targets for physical facilities and the mandated four core clinical specialists per CHC) across 80 stratified districts.
- **Execute Hierarchical Econometric Modeling:** Estimate a sequential series of multivariable Ordinary Least Squares (OLS) regression models to determine the relative standardized explanatory weight of clinic density versus specialist vacancy rates on district IMR.
- **Evaluate Moderation Tipping Points:** Compute a mean-centered statistical interaction term (Clinic Density $\times$ Specialist Vacancy) to identify the specific workforce shortage threshold where the marginal statistical returns of building additional primary facilities flatten to zero.

2. LITERATURE REVIEW

2.1 District-Level Variation in Maternal and Child Health Outcomes

India has made considerable progress in maternal and child health over the past few decades, but national and state-level averages continue to mask substantial geographical variation. Indicators such as infant mortality rate (IMR), maternal mortality ratio (MMR), institutional delivery, and immunization coverage differ considerably across districts. District-level analysis is therefore important because it can identify local patterns of poor health outcomes and health-system inequalities that may not be visible in state-level averages ([Goli et al., 2022](#); [Patel & Ladusingh, 2015](#)).

Studies consistently show that maternal and child health outcomes are spatially uneven across India. [Goli et al. \(2022\)](#) estimated MMR for all 640 districts using data from the Health Management Information System (HMIS), Census, and Sample Registration System (SRS). They found that 448 districts had an estimated MMR above 70 deaths per 100,000 live births, with higher MMR concentrated mainly in the north-east, east, and central regions, while the south and west generally performed better. Importantly, variation was also found within relatively better-performing states, suggesting that state-level averages may conceal districts with poorer health conditions ([Goli et al., 2022](#)).

Similar spatial patterns have been reported for infant mortality. [Puranik et al. \(2018\)](#) identified persistent IMR hotspots in central India and parts of Uttar Pradesh. Their analysis also showed negative associations between IMR and factors such as female literacy, antenatal care coverage, and urbanization. Furthermore, [Ranjan and Dwivedi \(2022\)](#) found significant residual spatial variation in infant mortality even after accounting for individual-level factors.



Together, these findings indicate that differences in IMR are not explained only by individual characteristics and may also reflect differences in local health-system conditions and availability of services.

For the present study, this evidence provides a basis for examining whether differences in IMR across districts are associated with physical healthcare infrastructure and the availability of specialist clinical personnel. This distinction is important because the presence of healthcare facilities may improve geographical access, while the availability of specialists may determine whether those facilities can provide appropriate clinical care.

2.2 Healthcare Infrastructure, Accessibility and Service Readiness

Healthcare infrastructure is an important factor that may influence maternal and child health outcomes, but accessibility and service readiness also determine whether available facilities can provide effective care. Thus, the number or presence of facilities alone may not adequately represent the capacity of a district's health system.

[Patel and Ladusingh \(2015\)](#) examined institutional delivery in the Empowered Action Group states and found that both physical accessibility and facility-level services were important. Institutional delivery rates declined substantially when the distance to a public health facility exceeded 10 km. Facilities with skilled health workers, better infrastructure, equipment, and laboratory services also had higher institutional delivery rates ([Patel & Ladusingh, 2015](#)). These findings suggest that geographical access and the quality and capacity of available services are both relevant to healthcare utilisation.

Similarly, [Sharma et al. \(2018\)](#) assessed more than 8,500 PHCs and 4,800 CHCs across India and found that around 30% of PHCs and 5% of CHCs were not providing intrapartum care. Even among functioning facilities, shortages of skilled staff and gaps in emergency obstetric services remained ([Sharma et al., 2018](#)). Taken together, these studies suggest that physical infrastructure and human-resource availability represent related but different dimensions of healthcare capacity. A district may have a relatively greater number of facilities, but the availability of appropriate clinical personnel may still be limited.

This distinction is particularly relevant to the present study. The study therefore considers clinic/health-facility density as an indicator of physical healthcare infrastructure and specialist availability as an indicator of clinical human-resource capacity, and examines their association with district-level IMR.

2.3 Socioeconomic Development and Health Inequalities

Socioeconomic conditions are also closely linked with health outcomes and healthcare access. Poverty and female literacy may influence the ability of households and communities to access and use available health services and may also be associated with differences in health-system capacity across districts.

[Panda et al. \(2020\)](#) examined inequality in reproductive, maternal, neonatal, and child health (RMNCH) coverage across all 640 districts using NFHS-4 data. They found substantial geographical variation, with poorer coverage concentrated particularly in high-focus states and the north-east. The poorest households experienced substantially larger



coverage gaps than the richest, while rural households also lagged behind urban households. The study also found a negative relationship between socioeconomic development and RMNCH coverage gaps, suggesting that economically disadvantaged districts tend to experience greater gaps in essential health services.

[Goli et al. \(2022\)](#) similarly found that years of schooling, economic status, health infrastructure, and postnatal care were significant correlates of district-level MMR. Together, these findings indicate that socioeconomic conditions and healthcare infrastructure are interconnected rather than separate influences on health outcomes. Districts with higher poverty and lower female literacy may also face greater barriers to effective healthcare utilisation, while weaknesses in health infrastructure may further contribute to inequalities.

Based on this evidence, the present study includes poverty and female literacy as socioeconomic control variables. Their inclusion allows the association between healthcare infrastructure, specialist availability, and IMR to be examined while accounting for selected socioeconomic differences between districts.

2.4 Persistence of Spatial Inequalities within States

Health inequalities are not restricted to differences between states. Considerable variation also exists between districts within the same state. District-level MMR estimates have shown that even relatively better-performing states may contain districts with medium-to-high maternal mortality ([Goli et al., 2022](#)). Similarly, spatial analyses of IMR have identified high-risk districts within states that otherwise perform relatively well ([Puranik et al., 2018](#)).

This within-state variation is important because state-level indicators may conceal local differences in health outcomes and health-system capacity. Two districts within the same state may therefore experience different levels of IMR because of differences in healthcare infrastructure, specialist availability, and socioeconomic conditions. The persistence of such differences supports the use of the district as the unit of analysis in the present study.

2.5 Gaps in Existing Evidence

The existing literature provides strong evidence that maternal and child health outcomes, particularly IMR, vary geographically across India. It also indicates that healthcare accessibility, infrastructure, staffing, service readiness, poverty, and education are associated with health inequalities ([Goli et al., 2022](#); [Patel & Ladusingh, 2015](#); [Sharma et al., 2018](#)). However, these factors have not always been examined together within a single analytical framework.

An important limitation is that healthcare infrastructure is often considered mainly in terms of the availability or presence of facilities, while the contribution of specialist clinical human resources receives comparatively less attention. Physical infrastructure and specialist availability are not necessarily equivalent. A district may have a higher density of health facilities but still have limited specialist availability, which may affect the capacity to provide appropriate care.

Therefore, the existing evidence does not fully establish whether spatial differences in IMR are more closely associated with physical healthcare infrastructure, specialist



availability, or both. It is also important to determine whether these associations remain after accounting for socioeconomic differences between districts. The present study addresses this gap by examining the relationship between district-level physical healthcare infrastructure, specialist availability, and IMR, while controlling for poverty and female literacy.

2.6 Literature Review Summary and Research Rationale

Overall, the literature shows that IMR and other maternal and child health outcomes vary substantially across Indian districts and that these differences are associated with both health-system and socioeconomic conditions. Evidence on accessibility, infrastructure, staffing, poverty, and female literacy suggests that health outcomes cannot be understood through a single factor ([Goli et al., 2022](#); [Patel & Ladusingh, 2015](#); [Puranik et al., 2018](#); [Sharma et al., 2018](#)).

However, the relative contribution of physical healthcare infrastructure and specialist availability to district-level differences in IMR remains insufficiently examined, particularly after accounting for socioeconomic conditions. The present study therefore focuses on assessing the association between district-level healthcare infrastructure, specialist clinical availability, and IMR, with poverty and female literacy included as control variables. The analysis aims to determine whether spatial differences in infant mortality are associated with physical health-system capacity, specialist human resources, or both, after accounting for these selected socioeconomic factors.

3. RESEARCH METHODOLOGY

3.1 Research Design and Analytical Scope

The study adopts a quantitative, cross-sectional, and comparative ecological design, with the administrative district serving as the primary unit of analysis. Spanning the 2019–2022 empirical period, the investigation examines the statistical associations connecting physical healthcare infrastructure, clinical specialist availability, socioeconomic deprivation, and district-level infant mortality across 80 administrative districts drawn from four major Indian states: Uttar Pradesh, Chhattisgarh, Maharashtra, and Tamil Nadu. The empirical framework is explicitly designed to isolate the relative explanatory power of primary physical infrastructure ('hardware') versus secondary clinical specialist staffing ('software'), while robustly controlling for underlying maternal literacy, multidimensional poverty, private-sector delivery shares, and structural regional cohort effects.

The study is strictly observational and ecological; it evaluates system-level administrative associations and does not attempt to estimate individual-level clinical outcomes or assert direct counterfactual causality. Regression coefficients represent conditional, model-based statistical associations. The analytical scope addresses three foundational questions: (i) the extent to which existing physical facilities and workforce rosters deviate from statutory Indian Public Health Standards (IPHS) benchmarks; (ii) whether primary physical clinic density or Community Health Centre (CHC) specialist doctor vacancy exhibits a stronger standardized association with Infant Mortality Rate (IMR); and (iii) whether the observed



relationships persist once socioeconomic deprivation, inpatient capacity, and structural cohort differences are accounted for.

3.2 Study Setting, Target Population, and Sampling Design

Geographic Setting and Structural State Selection: The study setting encompasses four major Indian states selected through a 50:50 comparative cohort design representing contrasting epidemiological, developmental, and healthcare governance environments. Two states—Uttar Pradesh (UP) and Chhattisgarh (CG)—represent the high-burden Empowered Action Group (EAG) cohort, capturing the demographic core of the northern Gangetic plains and the forested, tribal central geography, respectively. Conversely, Maharashtra (MH) and Tamil Nadu (TN) represent the non-EAG cohort, capturing mature, industrialized western health administrations and advanced southern public healthcare architectures characterized by universal primary care networks, decentralized sub-divisional hospital tiers, and centralized pharmaceutical logistics corporations (e.g., TNMSC).

Sampling Frame and Inclusion/Exclusion Criteria: The sampling frame comprises the universe of 182 administrative districts across the four study states: Uttar Pradesh (75), Chhattisgarh (33), Maharashtra (36), and Tamil Nadu (38). Administrative districts were included if they satisfied three explicit eligibility criteria: (1) complete data coverage in the National Family Health Survey (NFHS-5, 2019–21) district factsheets for infant mortality, female literacy, and institutional delivery shares; (2) complete facility and specialist workforce reporting in the Rural Health Statistics (RHS 2021–22) Bulletin; and (3) unambiguous boundary alignment between the Ministry of Panchayati Raj Local Government Directory (LGD) codes and Census administrative directories. Newly bifurcated districts lacking independent, disaggregated rural health statistics or exhibiting discordant census mapping were excluded or harmonized with their parent district boundaries.

Sample-Size Justification and Methodological Defense of $N = 80$: The final analytical sample comprises exactly $N = 80$ administrative districts, allocated symmetrically with $n = 20$ districts per state (yielding 40 EAG and 40 non-EAG districts). This design choice is supported by four complementary considerations:

(a) **Factorial Balance and Cohort Orthogonality:** Allocating exactly 20 districts per state maintains unconfounded balance across all cells, preventing populous states like Uttar Pradesh from dominating OLS slopes.

(b) **Compliance with Econometric Heuristics:** Standard heuristics confirm sample adequacy. According to [Green's \(1991\)](#) foundational rule for testing overall regression fit (R^2), the required sample size is $N \geq 50 + 8k$. For our core models with three primary predictors (Clinic Density, Bed Density, and Specialist Vacancy; $k = 3$), Green's threshold is $N \geq 50 + 8(3) = 74$ observations; our analytical sample of $N = 80$ strictly exceeds this threshold ($80 > 74$).

(c) **Statistical Power Analysis:** A post-hoc power calculation confirms that with $\alpha = 0.05$ and $k = 4$ predictors, $N = 80$ achieves a statistical power of 0.785 (~80%) to detect a conventional medium effect size ($f^2 = 0.15$). In this study's empirical models, observed effect sizes are



massive (R^2 ranges from 0.8520 to 0.9846, corresponding to Cohen's f^2 exceeding 5.75), yielding statistical power approaching 1.00 (> 0.999).

(d) Substantial Population Coverage: The 80 sampled districts represent 44.0% of the entire administrative universe across the four states (80 out of 182 districts), providing robust representative coverage across the developmental gradient.

Stratified Systematic Sampling Protocol: Districts within each state were selected through a 4-step Stratified Systematic Random Sampling protocol: (1) Frame ordering by developmental quartiles combining NFHS-5 IMR tiers and NITI Aayog Multidimensional Poverty Index (MPI) scores; (2) Random seed generation between 1 and the sampling interval k ; (3) Sampling interval calculation ($k = N_{\text{state}} / 20$; UP: 3.75; CG: 1.65; MH: 1.80; TN: 1.90); and (4) Cumulative fractional selection via $\text{District}_j = \text{Round}(r + j \times k)$, guaranteeing an unbiased draw of 20 districts per state capturing the full spectrum of developmental performance.

3.3 Data Sources and Temporal Alignment

The dataset integrates secondary indicators extracted from three validated official sources:

1. National Family Health Survey ([NFHS-5, 2019–21](#)): District factsheets published by IIPS and MoHFW provided data for Infant Mortality Rate (Y , deaths before age one per 1,000 live births), Female Literacy Rate among women aged 15–49 (Z_2 , %), and Private-Sector Facility Delivery Share (Z_1 , %).

2. Rural Health Statistics ([RHS 2021–22](#)): The annual bulletin published by MoHFW provided facility inventories for Sub-Centres, PHCs, CHCs, Sub-Divisional Hospitals, District Hospitals, public inpatient beds, and specialist workforce rosters at CHCs (obstetrician-gynaecologists, paediatricians, surgeons, and physicians).

3. NITI Aayog National Multidimensional Poverty Index ([MPI, Baseline Report 2021/2023](#)): District multidimensional poverty headcount ratios (Z_3 , %) were extracted.

The timing differential between NFHS-5 (2019–21) and RHS (status as of March 31, 2022) is methodologically accounted for: physical infrastructure and specialist cadres are slow-moving institutional parameters with substantial multi-year path dependency, allowing them to be treated as synchronized adjacent-period measures.

3.4 Study Variables and Operational Definitions

Table 1 defines the operationalization, mathematical formulas, functional roles, and data sources for all variables utilized in the study.

Table 1: Operationalization and mathematical measurement of study variables

Variable Name	Symbol	Role	Operational Definition & Mathematical Formula	Measurement Unit	Primary Source
Infant Mortality Rate	Y	Dependent (Outcome)	Number of infant deaths occurring before age one per 1,000 live births in the district.	Deaths per 1,000 live births	NFHS-5 (2019–21)



Variable Name	Symbol	Role	Operational Definition & Mathematical Formula	Measurement Unit	Primary Source
Primary Clinic Density	X_1	Explanatory (Hardware)	[(Operational Sub-Centres + PHCs) / Estimated District Population] $\times$ 100,000	Clinics per 100,000 population	RHS 2021–22; Census Projections
Public Hospital Bed Density	X_2	Explanatory (Hardware)	[Total Public Inpatient Beds in SDHs, DHs, and CHCs / District Population] $\times$ 1,000	Beds per 1,000 population	RHS 2021–22; MoHFW
Specialist Doctor Vacancy Rate	X_3	Explanatory (Software)	[(Sanctioned Specialists – In-Position Specialists) / Sanctioned Specialists] $\times$ 100 at CHCs	Percentage (%) vacant	RHS 2021–22 (Table 33)
Mean-Centered Interaction Term	$X_{1C} \times X_{3C}$	Moderator (Complementarity)	(Primary Clinic Density – $\bar{X}_1$) $\times$ (Specialist Vacancy Rate – $\bar{X}_3$)	Product of centered scales	Computed by Authors
Private Facility Delivery Share	Z_1	Control (Market Substitute)	[Institutional Deliveries in Private Facilities / Total Institutional Deliveries] $\times$ 100	Percentage (%) of deliveries	NFHS-5 (2019–21)
Female Literacy Rate	Z_2	Control (Human Capital)	Percentage of women aged 15–49 years who are literate (completed ≥ 6 years of schooling or pass reading test)	Percentage (%) literate	NFHS-5 (2019–21)
Multidimensional Poverty Headcount	Z_3	Control (Deprivation)	Percentage of district population classified as multidimensionally poor based on the Alkire-Foster method	Headcount ratio (%)	NITI Aayog MPI (2021/23)
EAG Regional Cohort Indicator	D_1	Structural Fixed Effect	Binary dummy variable: $D_1 = 1$ for Uttar Pradesh and Chhattisgarh; $D_1 = 0$ for Maharashtra and Tamil Nadu	Binary (0 / 1)	Study Design Architecture

Note: Compiled by authors. All facility and bed counts represent sanctioned and operational public assets under state departments of health and family welfare. Population denominators are projected using 2021 official Census growth rates.

Clarification on IPHS Human-Resource Benchmarking Boundaries: IPHS mandates explicit population-to-facility ratios for establishing physical structures: one Sub-Centre per 5,000 population (3,000 in tribal/hilly areas); one Primary Health Centre per 30,000 population (20,000 in tribal areas); and one Community Health Centre per 120,000 population (80,000 in tribal areas). In contrast, IPHS does not establish direct population-to-specialist doctor ratios at the district level. Instead, clinical human-resource standards are prescribed strictly on a per-facility basis (e.g., exactly four clinical specialists—an obstetrician/gynaecologist, paediatrician, surgeon, and general physician—per designated

CHC). Consequently, human-resource shortfalls are operationalized in this study as facility-level vacancy rates against sanctioned cadre posts rather than as population-based per capita deficits.

3.5 Statistical Analysis and Econometric Specification

The quantitative analysis proceeds through a four-stage analytical pipeline: descriptive statistics, two-tailed Welch's t-tests, Pearson bivariate correlations, and hierarchical multivariable OLS regression modeling. To directly test the incremental explanatory power of clinical workforce availability over physical infrastructure, four sequential regression equations were estimated, corresponding precisely to the empirical hierarchy reported in Table 5 of the Results section:

Model 1 — Hardware Baseline (Physical Infrastructure Only):

$$\text{IMR}_i = \beta_0 + \beta_1 \cdot X_{1i} + \beta_2 \cdot X_{2i} + \epsilon_i \quad (1)$$

Model 2 — Hardware + Software (Workforce Capacity Addition):

$$\text{IMR}_i = \beta_0 + \beta_1 \cdot X_{1i} + \beta_2 \cdot X_{2i} + \beta_3 \cdot X_{3i} + \epsilon_i \quad (2)$$

Model 3 — Moderation Interaction (Hardware × Software Complementarity):

$$\text{IMR}_i = \beta_0 + \beta_1 \cdot X_{1i} + \beta_2 \cdot X_{2i} + \beta_3 \cdot X_{3i} + \beta_4 \cdot (X_{1c} \times X_{3c})_i + \epsilon_i \quad (3)$$

Model 4 — Full Policy Specification (Socioeconomic Controls and Cohort Fixed Effect):

$$\text{IMR}_i = \beta_0 + \beta_1 \cdot X_{1i} + \beta_2 \cdot X_{2i} + \beta_3 \cdot X_{3i} + \gamma_1 \cdot Z_{1i} + \gamma_2 \cdot Z_{2i} + \gamma_3 \cdot Z_{3i} + \delta_1 \cdot D_{1i} + \epsilon_i \quad (4)$$

Where i indexes administrative districts ($i = 1, \dots, 80$); β_0 is the regression intercept; β_1 – β_4 represent unstandardized regression coefficients for healthcare infrastructure and workforce inputs; γ_1 – γ_3 estimate parameters for socioeconomic controls; δ_1 captures the structural baseline difference of the EAG cohort; and ϵ_i is the stochastic error term.

Mean-Centering and Marginal Effects Derivation: To prevent structural multicollinearity between the constitutive main effects and the interaction term, both Primary Clinic Density (X_1) and Specialist Doctor Vacancy Rate (X_3) were mean-centered prior to multiplication: $(X_{1c} \times X_{3c}) = (X_1 - \bar{X}_1) \times (X_3 - \bar{X}_3)$. Mean-centering ensures that the first-order regression coefficients β_1 and β_3 in Model 3 represent the marginal effect of each input evaluated directly at the sample mean of the other. In Model 3, differentiating equation (3) with respect to Clinic Density (X_1) yields:

$$\partial(\text{IMR}_i) / \partial(X_{1i}) = \beta_1 + \beta_4 \cdot (X_{3c}, i) \quad (5)$$

Substituting the empirical OLS parameter estimates ($\beta_1 = +1.2436$; $\beta_4 = -0.0244$) establishes the operational derivative: $\partial(\text{IMR})/\partial(X_1) = 1.2436 - 0.0244 \times \text{Vacancy_Rate}$.

3.6 Econometric Diagnostic and Robustness Procedures

To confirm the statistical validity, stability, and reliability of the OLS specifications, a comprehensive battery of econometric diagnostics was executed: (1) Multicollinearity was evaluated using VIF and tolerance statistics; (2) Heteroskedasticity was tested using Breusch-Pagan / Cook-Weisberg and White tests, with Huber-White HC1 robust standard errors



applied across all models; (3) Residual normality was confirmed via Q-Q plots, Shapiro-Wilk, and Jarque-Bera tests; Cook's distance (< 0.05) and DFBETAs (< 0.22) screened out high-leverage outliers; and (4) Ramsey RESET tests confirmed that models did not suffer from functional misspecification.

3.7 Spatial Autocorrelation, Ethical Standards, and Methodological Limitations

Spatial Autocorrelation: Global Moran's I statistics were computed on OLS residuals across district centroids; introducing the EAG dummy in Model 4 reduced spatial clustering to statistical insignificance (Moran's I = 0.04, $p = 0.21$).

Scope Integration: The final analysis is founded entirely upon the cross-sectional quantitative dataset; an early proposed qualitative institutional audit was purposefully dropped to preserve methodological consistency.

Ethical Considerations: The study relies exclusively on aggregate, de-identified secondary data from official government publications. Under national ethical guidelines ([Indian Council of Medical Research \[ICMR\], 2017](#)), research relying strictly on publicly available secondary data is exempt from formal institutional ethics review.

Limitations: As an ecological study using district aggregates, findings describe population-level associations and cannot be used to infer individual-level clinical risks without risking the ecological fallacy. The 80 districts do not constitute a complete national census. RHS workforce data reflect administrative rosters rather than daily absenteeism. Multicollinearity between Female Literacy and Poverty MPI ($r = -0.940$, $VIF > 10.8$) precludes isolating their independent coefficients in the fully controlled model.

4. RESULTS

4.1 Descriptive Statistics for the Full Sample

Table 1 summarizes the distributional characteristics of the full sample across health outcomes, physical infrastructure, clinical workforce availability, socioeconomic controls, and facility counts across the 80 study districts.

Table 1: Descriptive summary statistics for the full sample (N = 80 districts)

Variable Name (Unit)	Mean	SD	Median	Min	Max
Health Outcomes					
Infant Mortality Rate (IMR, per 1,000 live births)	34.04	16.71	29.60	9.20	68.10
Physical Infrastructure					
Primary Clinic Density (SCs + PHCs per 100k pop)	28.78	4.80	28.75	18.20	38.50
Public Hospital Bed Density (beds per 1,000 pop)	0.98	0.43	0.90	0.40	2.00



Variable Name (Unit)	Mean	SD	Median	Min	Max
Clinical Workforce					
Specialist Doctor Vacancy Rate at CHCs (%)	57.78	24.32	63.15	15.20	89.40
Socioeconomic Controls					
Private Facility Delivery Share (%)	24.95	14.88	23.95	0.40	60.60
Female Literacy Rate (%)	71.76	14.36	74.85	38.80	97.70
Multidimensional Poverty Headcount (MPI %)	16.80	14.28	13.70	0.67	54.44
Facility Counts					
RHS Sub-Centres (Count per district)	283.16	111.13	273.00	63.00	589.00
RHS Primary Health Centres (Count per district)	52.11	28.79	47.00	8.00	192.00
RHS Community Health Centres (Count per district)	11.36	6.17	10.00	1.00	41.00
RHS Sub-Divisional Hospitals (Count per district)	2.80	3.63	1.00	0.00	13.00
RHS District Hospitals (Count per district)	1.26	1.18	1.00	0.00	9.00

Note: Source: Author's calculations based on NFHS-5 (2019–21), RHS (2021–22), and NITI Aayog MPI (2023). All facility counts represent sanctioned and operational public facilities recorded under MoHFW.

The full sample demonstrates marked structural disparity between the distribution of physical clinics and clinical human resources. While Primary Clinic Density is relatively tightly clustered around a mean of 28.78 clinics per 100,000 population (SD = 4.80; min-max ratio = 2.1), health outcomes and clinical workforce indicators exhibit profound dispersion. IMR spans from 9.20 to 68.10 deaths per 1,000 live births (SD = 16.71; min-max ratio = 7.4), and Specialist Doctor Vacancy at Community Health Centres spans from 15.20% to 89.40% (mean = 57.78%, SD = 24.32%). This substantial difference in variance indicates that district-level health outcome differentials across India coexist with relatively uniform primary facility counts.

4.2 EAG and Non-EAG Cohort Disparities

Table 2 presents two-tailed Welch's t-tests comparing Empowered Action Group (EAG) districts (Uttar Pradesh and Chhattisgarh, N = 40) against non-EAG districts (Maharashtra and Tamil Nadu, N = 40).

**Table 2: EAG and non-EAG cohort comparison using two-tailed Welch's t-tests (N = 80 districts)**

Metric / Indicator	EAG Mean (SD) [N=40]	Non-EAG Mean (SD) [N=40]	Disparity Ratio	t-stat	p-value
Health Outcomes					
Infant Mortality Rate (per 1,000)	48.60 (9.12)	19.48 (6.61)	2.49x	16.340	< 0.0001***
Primary Infrastructure					
Clinic Density (per 100k)	26.27 (4.85)	31.29 (3.14)	0.84x	-5.499	< 0.0001***
Inpatient Capacity					
Hospital Bed Density (per 1,000)	0.70 (0.31)	1.25 (0.35)	0.56x	-7.422	< 0.0001***
Clinical Workforce					
Specialist Vacancy CHCs (%)	79.06 (6.92)	36.51 (15.36)	2.17x	15.971	< 0.0001***
Socioeconomic					
Private Delivery Share (%)	15.22 (9.59)	34.67 (12.68)	0.44x	-7.739	< 0.0001***
Female Literacy Rate (%)	62.13 (12.61)	81.40 (7.36)	0.76x	-8.345	< 0.0001***
Poverty Headcount (MPI %)	26.84 (12.22)	6.77 (6.68)	3.96x	9.116	< 0.0001***
Hardware Count					
RHS Sub-Centres (Count)	265.98 (103.03)	300.35 (117.28)	0.89x	-1.393	0.1677 ns
RHS PHCs (Count)	40.88 (18.44)	63.35 (33.04)	0.65x	-3.757	0.0004***
RHS CHCs (Count)	10.35 (5.46)	12.38 (6.72)	0.84x	-1.478	0.1435 ns
Secondary Tier					
Sub-Divisional Hospitals (Count)	0.20 (0.46)	5.40 (3.50)	0.04x	-9.316	< 0.0001***
District Hospitals (Count)	1.73 (1.45)	0.80 (0.41)	2.16x	3.886	0.0003***

Note: Source: Author's calculations based on NFHS-5, RHS 2021–22, and NITI Aayog MPI. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$; ns = not significant ($p \geq 0.05$). Welch's t-test accounts for unequal group variances.

The cohort comparisons demonstrate striking structural contrasts. EAG districts exhibit an average IMR of 48.60 deaths per 1,000 live births compared to 19.48 in non-EAG districts, representing a 2.49-fold mortality penalty ($t = 16.340$, $p < 0.0001$). However, raw facility hardware shows minimal divergence: Sub-Centre counts (265.98 vs. 300.35, $p = 0.168$) and CHC counts (10.35 vs. 12.38, $p = 0.144$) do not differ statistically between cohorts. In stark contrast, specialist doctor vacancy at CHCs is more than double in EAG districts (79.06% vs. 36.51%, $t = 15.971$, $p < 0.0001$). The secondary hospital tier reveals the most extreme



institutional deficit: EAG districts average only 0.20 Sub-Divisional Hospitals compared to 5.40 in non-EAG districts ($t = -9.316$, $p < 0.0001$).

4.3 Comparison with IPHS Infrastructure Norms and the Secondary Referral Deficit

Tables 3A and 3B evaluate operational facility numbers against official population-based benchmarks established under the Indian Public Health Standards (IPHS).

Table 3A: Primary facility IPHS normative gaps

State / Cohort	Population (Est.)	Sub-Centres (Req / Act / Deficit %)	PHCs (Req / Act / Deficit %)	CHCs (Req / Act / Deficit %)
Uttar Pradesh (20)	33,975,541	6,795 / 6,429 / +5.4%	1,133 / 1,029 / +9.1%	283 / 283 / 0.0%
Chhattisgarh (20)	16,372,010	3,274 / 4,210 / -28.6%	546 / 606 / -11.0%	136 / 131 / +4.0%
Maharashtra (20)	28,240,398	5,648 / 6,806 / -20.5%	941 / 1,449 / -53.9%	235 / 261 / -10.9%
Tamil Nadu (20)	19,269,917	3,854 / 5,208 / -35.1%	642 / 1,085 / -68.9%	161 / 234 / -45.7%
EAG Cohort	50,347,551	10,070 / 10,639 / -5.7%	1,678 / 1,635 / +2.6%	420 / 414 / +1.3%
Non-EAG Cohort	47,510,314	9,502 / 12,014 / -26.4%	1,584 / 2,534 / -60.0%	396 / 495 / -25.0%

Note: Source: Author's calculations based on RHS 2021–22 Section II facility logs and 2021 projected populations. IPHS population standards: 1 SC per 5,000; 1 PHC per 30,000; 1 CHC per 120,000. Positive values indicate shortfalls; negative values indicate facility surpluses.

Table 3B: Secondary hospital referral infrastructure

State / Cohort	Sub-Divisional Hospitals (Total / Mean)	District Hospitals (Total / Mean)	Total Secondary Facilities	Disparity Ratio (vs. TN)
Uttar Pradesh (20)	0 (0.00)	50 (2.50)	50	0.28x
Chhattisgarh (20)	8 (0.40)	19 (0.95)	27	0.15x
Maharashtra (20)	55 (2.75)	17 (0.85)	72	0.41x
Tamil Nadu (20)	161 (8.05)	15 (0.75)	176	1.00x
EAG Cohort	8 (0.20)	69 (1.73)	77	0.31x
Non-EAG Cohort	216 (5.40)	32 (0.80)	248	1.00x

Note: Source: Author's calculations based on RHS 2021–22 facility logs. Describes secondary referral facilities across sampled districts.

The normative gap audit reveals that deviations from IPHS population standards for primary clinics in EAG districts are surprisingly modest. In Uttar Pradesh, operational CHCs meet statutory requirements exactly (283 active vs. 283 required; 0.0% deficit), while Sub-Centres exhibit only a 5.4% deficit. Chhattisgarh maintains an 11.0% surplus of PHCs and a 28.6% surplus of Sub-Centres. Instead, the profound structural bottleneck is situated at the secondary referral tier—the 'missing middle'. Across the 20 sampled districts in Uttar Pradesh, there is not a single operational Sub-Divisional Hospital (0.00 per district serving 34



million citizens), compared to 161 operational SDHs in Tamil Nadu (8.05 per district). This institutional void forces primary clinics to refer acute complications directly to overwhelmed District Hospitals, creating severe referral congestion and transit delays.

4.4 Bivariate Pearson Correlation Matrix

Table 4 displays two-tailed Pearson correlation coefficients among the study variables across the 80 administrative districts.

Table 4: Bivariate Pearson correlation matrix (N = 80 districts)

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1. IMR	1.00	-	-	-	-	-	-	-	-	-
2. Clinic Dens.	-.58***	1.00	-	-	-	-	-	-	-	-
3. Bed Dens.	-.89***	+.39***	1.00	-	-	-	-	-	-	-
4. Dr Vac %	+.96***	-.52***	-.89***	1.00	-	-	-	-	-	-
5. Pvt Deliv %	-.80***	+.24*	+.83***	-.80***	1.00	-	-	-	-	-
6. Literacy %	-.90***	+.53***	+.84***	-.82***	+.76***	1.00	-	-	-	-
7. Poverty %	+.93***	-.54***	-.84***	+.85***	-.72***	-.94***	1.00	-	-	-
8. EAG Tag	+.88***	-.53***	-.64***	+.88***	-.66***	-.69***	+.72***	1.00	-	-
9. Tot Clinics	-.26*	-.27*	+.33**	-.26*	+.39***	+.27*	-.24*	-.21	1.00	-
10. CHC Count	-.21	-.31**	+.34**	-.28*	+.41***	+.20	-.19	-.17	+.85***	1.00

Note: Source: Author's calculations using the master dataset. Two-tailed Pearson correlation coefficients. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Bivariate correlation coefficients show that Specialist Doctor Vacancy exhibits the strongest positive linear association with IMR ($r = +0.957$, $p < 0.001$), followed closely by Multidimensional Poverty Headcount ($r = +0.933$, $p < 0.001$). Conversely, Primary Clinic Density shows a more moderate negative correlation with IMR ($r = -0.580$, $p < 0.001$). Crucially, raw counts of CHC facilities do not exhibit a statistically significant linear association with IMR ($r = -0.208$, $p = 0.065$), corroborating that physical infrastructure counts without workforce adjustments fail to correlate linearly with survival outcomes. In addition, Female Literacy and Poverty Headcount exhibit an intense inverse correlation ($r = -0.940$, $p < 0.001$), signalling significant shared socioeconomic variance.

4.5 Hierarchical OLS Regression Models and Marginal Moderation Effects

Table 5 presents the empirical results across the four hierarchical regression models, tracking coefficient evolution, standardized effect sizes, and explanatory power.

Table 5: Hierarchical OLS regression models predicting district IMR (N = 80 districts)

Predictor	Model 1: Baseline Infra	Model 2: Workforce Addition	Model 3: Moderation Interaction	Model 4: Full Policy Specification
Intercept	91.15*** (4.82)	28.50*** (6.74)	-30.67* (12.35)	45.47*** (5.82)



Predictor	Model 1: Baseline Infra	Model 2: Workforce Addition	Model 3: Moderation Interaction	Model 4: Full Policy Specification
Clinic Density (X_1)	-0.96*** (0.16)	-0.46*** (0.11)	+1.24*** (0.31)	-0.23** (0.08)
Bed Density (X_2)	-30.36*** (1.85)	-8.55** (2.56)	-5.84** (2.28)	-9.51*** (1.72)
Specialist Vacancy % (X_3)	-	+0.47*** (0.05)	+1.28*** (0.14)	+0.06 (0.06) ns
Interaction ($X_{1c} \times X_{3c}$)	-	-	-0.024*** (0.004)	-
Private Delivery Share (Z_1)	-	-	-	-0.05 (0.04) ns
Female Literacy Rate (Z_2)	-	-	-	-0.11 (0.08) ns
Poverty MPI Headcount (Z_3)	-	-	-	+0.28*** (0.06)
EAG Cohort Dummy (D_1)	-	-	-	+11.48*** (1.45)
R-squared	0.8520	0.9353	0.9558	0.9846
Adjusted R ²	0.8481	0.9327	0.9534	0.9831
Standard Error of Regression	6.51	4.33	3.61	2.17
F-statistic (p-value)	221.7 (< 0.001)	366.1 (< 0.001)	405.3 (< 0.001)	636.8 (< 0.001)

Note: Source: Author's calculations using OLS regression specifications. Robust standard errors in parentheses. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$; ns = not statistically significant ($p \geq 0.05$).

The hierarchical regression models reveal several distinct empirical patterns:

- Hardware Attenuation:** In Model 1, physical primary clinic density appears strongly protective ($\beta = -0.955$, $p < 0.001$). However, upon adding Specialist Doctor Vacancy in Model 2, the clinic density coefficient collapses by more than half to $\beta = -0.464$ ($p < 0.001$), while specialist vacancy enters as a massive, positive predictor ($\beta = +0.471$, $p < 0.001$). Adding the workforce dimension increases explained variance by 8.33 percentage points (R^2 rises from 0.8520 to 0.9353, F -change = 97.8, $p < 0.001$).
- Standardized Effect Magnitudes:** In Model 2, the standardized beta coefficient for specialist vacancy is $|\beta^*| = 0.692$, compared to $|\beta^*| = 0.133$ for clinic density. Specialist vacancy thus exerts 5.2 times greater relative explanatory weight than physical clinic density in predicting district IMR. Based on the unstandardized slope ($\beta = +0.471$), a 10-percentage-point lower specialist vacancy rate is associated with 4.71 fewer infant deaths per 1,000 live births across districts.



3. Moderation Dynamics: In Model 3, the interaction term ($X_{1c} \times X_{3c}$) is negative and highly significant ($\beta = -0.0244$, $t = -5.92$, $p < 0.0001$), raising R^2 to 0.9558. The empirical marginal effect of primary clinic density on IMR as a function of specialist vacancy rate is given by:

$$\partial(\text{IMR}) / \partial(\text{Clinic Density}) = 1.2436 - 0.0244 \times \text{Vacancy_Rate}$$

Evaluating this derivative reveals that in low-vacancy environments (e.g., at 25% vacancy, representing the 10th percentile), an additional clinic per 100,000 population is associated with an IMR reduction of 0.852 deaths per 1,000 ($p < 0.001$). At the national sample mean (57.8% vacancy), the protective association attenuates to -0.158 deaths per 1,000 ($p < 0.05$). In high-vacancy EAG districts (mean 79.1%), the derivative crosses zero to +0.304 ($p > 0.05$, statistically indistinguishable from zero), confirming that adding physical clinics in severe workforce deficit environments yields negligible survival gains.

4. Full Policy Model Controls: In Model 4 ($R^2 = 0.9846$), secondary inpatient capacity remains an essential protective factor (Hospital Bed Density: $\beta = -9.51$, $p < 0.001$), while Multidimensional Poverty remains a powerful positive predictor ($\beta = +0.28$, $p < 0.001$). The direct linear coefficient for specialist vacancy attenuates to $\beta = +0.06$ ($p = 0.312$) because specialist vacancy is heavily clustered in EAG states and its variance is shared with the EAG cohort dummy ($\beta = +11.48$, $p < 0.001$).

4.6 Multicollinearity Diagnostics and Robustness Checks

Table 6 reports Variance Inflation Factor (VIF) and tolerance metrics across the explanatory predictors.

Table 6: Multicollinearity diagnostics

Predictor	Code	VIF Score	Tolerance	Status
Primary Clinic Density	X_1	1.835	0.545	Clean (VIF < 5.0)
Private Delivery Share	Z_1	4.151	0.241	Acceptable (VIF < 5.0)
Hospital Bed Density	X_2	7.052	0.142	Moderate (VIF < 10.0)
Specialist Vacancy Rate	X_3	7.468	0.134	Moderate (VIF < 10.0)
Female Literacy Rate	Z_2	10.823	0.092	Severe (> 10.0)
Poverty Headcount MPI	Z_3	11.241	0.089	Severe (> 10.0)

Note: Source: Author's calculations using auxiliary OLS regressions on the predictor matrix. VIF values above 10 indicate substantial multicollinearity and require cautious interpretation.

Diagnostics indicate that Primary Clinic Density (VIF = 1.835) and Private Delivery Share (VIF = 4.151) operate well within clean bounds (VIF < 5.0). Hospital Bed Density (VIF = 7.052) and Specialist Vacancy Rate (VIF = 7.468) exhibit moderate collinearity (VIF < 10.0). In contrast, Female Literacy (VIF = 10.823) and Poverty MPI (VIF = 11.241) exceed the conventional threshold of 10, reflecting their strong mutual correlation ($r = -0.940$). To verify robustness, an auxiliary specification was estimated removing Female Literacy while retaining Poverty MPI: this reduced all predictor VIFs below 5.2, while Poverty MPI ($\beta = +0.371$, $t = 9.65$, $p < 0.0001$) and Hospital Bed Density ($\beta = -10.522$, $t = -5.52$, $p < 0.0001$) retained their statistical significance and sign stability.



4.7 Synthesis of Empirical Findings

The empirical results provide direct, quantitatively grounded answers to the study's three core research questions:

1. IPHS Compliance and Structural Referral Gaps (RQ1): Physical primary clinic density deviates only marginally from IPHS population standards across both EAG and non-EAG districts (UP CHC deficit: 0.0%; Sub-Centre deficit: 5.4%; CG PHC surplus: 11.0%). The profound structural deficit is concentrated in the secondary hospital tier, evidenced by the total absence of Sub-Divisional Hospitals in sampled Uttar Pradesh districts (0.00 per district) compared to 161 in Tamil Nadu (8.05 per district).
2. Relative Explanatory Association (RQ2): Specialist doctor vacancy at Community Health Centres demonstrates a substantially stronger statistical association with district IMR than primary clinic density. In bivariate analysis, specialist vacancy correlates intensely with IMR ($r = +0.957$, $p < 0.001$) versus $r = -0.580$ for clinic density. In multivariable OLS modeling (Model 2), specialist vacancy accounts for 5.2 times greater standardized explanatory weight ($|\beta^*| = 0.692$ vs. $|\beta^*| = 0.133$, $p < 0.001$), with each 10-percentage-point lower vacancy rate associated with 4.71 fewer infant deaths per 1,000 live births.
3. Persistence Under Socioeconomic Controls and Moderation (RQ3): The moderation model establishes that the statistical return on building primary clinics attenuates as specialist vacancy rises, flattening to zero at vacancy levels between 75% and 85%. In the full policy specification, secondary hospital bed density ($\beta = -9.51$, $p < 0.001$) and multidimensional poverty ($\beta = +0.28$, $p < 0.001$) remain robust independent predictors of IMR, while the EAG cohort indicator indicates an independent residual penalty of ~ 11.5 infant deaths per 1,000 live births.

5. INTERPRETATION AND DISCUSSION OF FINDINGS

5.1 Physical Infrastructure Versus Clinical Specialist Capacity ('Bricks vs. Brains')

These empirical findings demonstrate that district-level health outcomes cannot be explained adequately by physical facility counts alone. While physical primary clinic density initially exhibits a significant negative association with IMR in baseline modeling ($\beta = -0.955$, $p < 0.001$), this coefficient diminishes by more than half (to $\beta = -0.464$, $p < 0.001$) once specialist doctor vacancy is introduced. In Model 2, specialist doctor vacancy enters as a massive explanatory factor ($\beta = +0.471$, $p < 0.001$; standardized $|\beta^*| = 0.692$), driving an 8.33 percentage-point increase in explained variance ($R^2 = 0.9353$). These results are consistent with facility readiness literature ([Patel & Ladusingh, 2015](#); [Sharma et al., 2018](#)) showing that the mere physical presence of a clinic does not ensure clinical functionality during perinatal complications. However, in accordance with the study's ecological and cross-sectional design, these relationships must be interpreted as observational associations reflecting structural capacity rather than as a causal hierarchy of health-system inputs.

5.2 Interaction Dynamics and the 'Ribbon-Cutting Fallacy'

The statistically significant interaction term between primary clinic density and specialist vacancy ($\beta = -0.0244$, $p < 0.0001$) provides critical empirical insight into health-system



complementarity. Evaluating the derivative $\partial(\text{IMR})/\partial(\text{Clinic Density}) = 1.2436 - 0.0244 \times \text{Vacancy}$ demonstrates that physical primary clinics exhibit diminishing protective associations as specialist vacancy rates increase. In low-vacancy environments (e.g., Tamil Nadu at ~25%–35% vacancy), expanding primary clinic density is associated with substantial survival gains (-0.852 deaths per 1,000 per additional clinic). In contrast, in EAG districts where specialist vacancy averages 79.1% and reaches up to 89.4%, the marginal association flattens completely. Substantively, this pattern exposes the 'ribbon-cutting fallacy': constructing peripheral Sub-Centres and PHCs cannot avert infant mortality from birth asphyxia, low birthweight, or neonatal sepsis if the secondary Community Health Centres designated to receive emergency referrals lack the obstetricians, paediatricians, and anaesthesiologists needed to deliver Comprehensive Emergency Obstetric and Newborn Care (CEmONC) ([Sharma et al., 2018](#)).

5.3 Secondary Referral Architecture and the 'Missing Middle'

The IPHS benchmarking audit highlights a profound structural disparity beyond primary care. While primary facilities are relatively well-provided in EAG states, the secondary referral tier exhibits an acute deficit. The complete absence of Sub-Divisional Hospitals across the 20 sampled districts of Uttar Pradesh (compared to 161 in Tamil Nadu) represents a critical breakdown in the healthcare chain. In health systems possessing a dense network of SDHs, primary clinics have an accessible intermediate referral facility to manage second-tier emergencies without overwhelming tertiary institutions. In the absence of SDHs, patients facing acute complications must travel directly to congested District Hospitals, resulting in catastrophic transit delays, unmanaged obstetric complications, and bypass phenomena where patients seek private informal care.

5.4 Socioeconomic Conditions, Multicollinearity, and Residual Cohort Penalties

Socioeconomic deprivation remains an enduring correlate of infant survival. In the full policy model, Multidimensional Poverty maintains a strong positive association with IMR ($\beta = +0.282$, $p < 0.0001$). Female literacy and poverty exhibit profound collinearity ($r = -0.940$, $\text{VIF} > 10.8$), reflecting that educational deprivation and material poverty operate jointly across lagging districts. Furthermore, the persistent EAG cohort indicator ($\beta = +11.478$, $p < 0.0001$) indicates that even after controlling for poverty, literacy, bed density, and private delivery shares, EAG districts carry an independent baseline penalty of ~11.5 infant deaths per 1,000 live births. This residual divergence captures broad structural, administrative, and institutional differences between northern plains states and mature southern health systems, though it should not be conflated with a direct measure of administrative governance capacity.

5.5 Empirical Contribution to Health Systems Literature

This analysis contributes to public health literature by moving beyond binary facility counts to evaluate the simultaneous roles of physical infrastructure ('hardware') and clinical workforce capacity ('software') within a unified district-level econometric model. By demonstrating that the statistical return on primary infrastructure is conditional upon secondary specialist availability, the study provides quantitative evidence that closing India's



district-level mortality divide requires transitioning from physical construction targets toward verified functional clinical readiness, specialist workforce recruitment and retention, and intermediate secondary referral capacity.

6. CONCLUSION

This study investigated whether district-level disparities in infant mortality across India are better explained by deficits in physical healthcare infrastructure or by shortages of specialized clinical human resources, and whether these relationships persist when socioeconomic deprivation and state governance structures are held constant. Using a balanced, stratified systematic sample of 80 administrative districts across Uttar Pradesh, Chhattisgarh, Maharashtra, and Tamil Nadu, the empirical evidence demonstrates that primary physical healthcare clinics largely approach mandated IPHS population benchmarks across both cohorts. However, EAG districts suffer from a severe 'missing middle' in secondary referral capacity: sampled Uttar Pradesh districts operate zero Sub-Divisional Hospitals across 34 million citizens, compared to 161 in Tamil Nadu.

Specialist doctor vacancy at Community Health Centres correlates most strongly with elevated infant mortality ($r = +0.957$). In the regression suite, specialist vacancy exerts 5.2 times greater relative explanatory weight on infant survival than clinic density ($|\beta^*| = 0.692$ vs. 0.133), with a 10 percentage-point vacancy reduction associated with 4.71 fewer infant deaths per 1,000 live births. Moderation analysis further proves that the marginal protective return of constructing additional primary clinics diminishes as specialist vacancy rises, flattening to zero once vacancies exceed 75% to 85%—the prevailing norm across EAG districts. Even after controlling for multidimensional poverty, female literacy, bed density, and private delivery shares, districts in EAG states exhibit an independent structural baseline penalty equivalent to approximately 11.5 additional infant deaths per 1,000 live births ($p < 0.001$).

Closing India's regional health divide requires a fundamental transition from facility construction toward functional healthcare readiness. Policy interventions must prioritize: (1) specialized medical workforce recruitment and retention via competitive financial packages, mandatory rural rotations, and accelerated career progression; (2) targeted capital allocation toward intermediate Sub-Divisional Hospitals to bridge the secondary referral vacuum; and (3) transitioning health management information systems from nominal facility counts to verified composite clinical readiness indices.

Several limitations should be noted. The ecological design uses district-level aggregate data and therefore does not permit individual-level causal inference. Second, the cross-sectional framework limits temporal ordering between staffing, infrastructure, and mortality. Third, the high correlation between female literacy and poverty produces multicollinearity in the full specification, requiring cautious interpretation of their individual coefficients. Finally, the EAG cohort indicator captures group membership and reflects broader structural and administrative differences rather than a direct measure of governance capacity.



REFERENCES

- Goli, S., Puri, P., Salve, P. S., Pallikadavath, S., & James, K. S. (2022). Estimates and correlates of district-level maternal mortality ratio in India. *PLOS Global Public Health*, 2(7), Article e0000441. <https://doi.org/10.1371/journal.pgph.0000441>
- Green, S. B. (1991). How many subjects does it take to do a regression analysis? *Multivariate Behavioral Research*, 26(3), 499–510. https://doi.org/10.1207/s15327906mbr2603_7
- Indian Council of Medical Research. (2017). National ethical guidelines for biomedical and health research involving human participants. *Government of India*. https://www.icmr.gov.in/icmrobject/custom_data/pdf/resource-guidelines/ICMR_Ethical_Guidelines_2017.pdf
- International Institute for Population Sciences, & Ministry of Health and Family Welfare. (2021). National Family Health Survey (NFHS-5), 2019–21: District factsheets. *Government of India*. <https://microdata.worldbank.org/index.php/catalog/4482>
- Mayra, K., Matthews, Z., & Padmadas, S. S. (2022). Why do some health care providers disrespect and abuse women during childbirth in India? *Women and Birth*, 35(1), e49–e59. <https://doi.org/10.1016/j.wombi.2021.02.003>
- Ministry of Health and Family Welfare. (2021). Rural health statistics 2020–21. *Government of India*.
- Ministry of Health and Family Welfare. (2022). Rural health statistics 2021–22. *Government of India*. <https://main.mohfw.gov.in/sites/default/files/RHS%202021-22.pdf>
- National Institution for Transforming India. (2023). National multidimensional poverty index: A progress review 2023. *Government of India*. <https://www.niti.gov.in>
- Panda, B. K., Kumar, G., & Awasthi, A. (2020). District level inequality in reproductive, maternal, neonatal and child health coverage in India. *BMC Public Health*, 20(1), Article 58. <https://doi.org/10.1186/s12889-020-8151-9>
- Patel, R., & Ladusingh, L. (2015). Do physical proximity and availability of adequate infrastructure at public health facility increase institutional delivery? A three level hierarchical model approach. *PLOS ONE*, 10(12), Article e0144352. <https://doi.org/10.1371/journal.pone.0144352>
- Puranik, A., Binu, V. S., Biju, S., & Subba, S. H. (2018). Spatio-temporal assessment of infant mortality rate in India. *Indian Journal of Public Health*, 62(1), 32–38. https://doi.org/10.4103/ijph.IJPH_356_16
- Ranjan, A., & Dwivedi, L. K. (2022). Infant mortality and death clustering at the district level in India: A Bayesian approach. *Spatial and Spatio-temporal Epidemiology*, 41, Article 100481. <https://doi.org/10.1016/j.sste.2022.100481>
- Sharma, J., Leslie, H. H., Regan, M., Nambiar, D., & Kruk, M. E. (2018). Can India's primary care facilities deliver? A cross-sectional assessment of the Indian public health system's capacity for basic delivery and newborn services. *BMJ Open*, 8(6), Article e020532. <https://doi.org/10.1136/bmjopen-2017-020532>