A space-time-age model for subnational child mortality estimation in India

Nathaniel Henry

1   Introduction

In 2017, the Indian Ministry of Health and Family Welfare released the National Health Policy (NHP), the first strategic plan in 15 years to clearly lay out the Indian government's priorities and targets related to health.[1] This document's importance extends even beyond
5   its goal-setting function for the next ten years of health policy in India: it also represents an attempt by India's national governing party to fulfill campaign promises centered around universal health coverage by the year 2025.[2] To carry out these core principles, Indian health policymakers first require local and focal insights about health burden nationwide. A situation analysis released alongside the NHP emphasizes how measuring inequality and
10   diversity in health outcomes is a first step towards achieving universal good health:

> We also need to keep in mind that high degree of inequity in health outcomes
> and access to health care services exists in India. This is evidenced by indicators
> disaggregated for vulnerable groups and between and within States. Identify-
> ing the deprived areas/vulnerable population groups (including special groups)
15   > through disaggregated data is a first step to address the existing inequities in
> health outcomes between and within States in India.[3]

The NHP places particular emphasis on the health and survival of one vulnerable sub-population: children under five years of age. The NHP's concrete goals for the coming decade include commitments to reduce the neonatal mortality rate (NMR) to fewer than 16
20   deaths per 1,000 live births by 2025; to reduce the infant mortality rate (IMR) to fewer than 28 deaths per 1,000 live births by 2018; and to reduce the under-5 mortality rate (U5MR) to fewer than 23 deaths per 1,000 live births by 2025. To meet these targets universally and equitably, policymakers need information about disparities in the health of children under 1 month, 1 year, and 5 years of age, respectively, across the country. However, none of India's
25   three primary mortality surveillance systems have traditionally offered spatially-resolved information about child survival across the country.

In this chapter, I describe a model to estimate neonatal, infant, and child mortality at the district level based on household survey data, the only data source for child mortality for which location data is published below the state level. Based on results from this model, I demonstrate how district-level estimates of mortality can reveal policy-relevant information hidden by state-level results. I then compare these estimates to the data available from the other two mortality surveillance systems to demonstrate how subnational reporting across all three systems is needed to support equitable child health outcomes across India.

1.1 Mortality across Indian states: progress, transition, and inequality

The NHP's emphases on child welfare and equity reflect India's past experience in child health provision. While India halved the under-5 mortality rate from 2000 to 2017, from 80 deaths per 1,000 live births to less than 40,[4] state-level estimates of child welfare indicate that some parts of the country are being left behind amid general progress towards improved child health. The NHP reports that as of 2013, the states of Madhya Pradesh and Assam both experienced infant mortality rates of 54/1,000, or more than 1 in 20, more than five times higher than the IMR observed across the country in Goa (9/1,000) or Manipur (10/1,000).[1]

Previous research has contextualized these striking disparities as part of a discontinuity in the standard epidemiological and demographic transitions. In some regions of India, such as Kerala state in the south, life expectancy has increased dramatically, leading to an increase in non-communicable disease burden and corresponding strain on the health system related to elder care;[5] meanwhile, particularly in rural settings and among marginalized groups, infectious diseases and maternal and child disorders are more deleterious to health.[6] These competing needs complicate national health policymaking, leading some researchers to contend that the Indian epidemiological context must be understood as many "nations within a nation."[6]

Facing a rapidly-changing health context marked by fundamental differences in health needs, Indian policymakers need health information systems that accurately reflect local variation in health status in order to deploy appropriate interventions within communities facing the greatest disease burden. The World Health Organization recognizes health information systems as a necessary building block for any successful health system, following the principle that public health relies on evidence to function.[7,8] Do India's health information systems live up to the National Health Policy's commitments on child mortality? This chapter explores whether

1.2 Tracking local variation in mortality across India

According to the 1969 Registration of Births and Deaths Act passed by the Parliament of India, birth and death registration are owed to every Indian citizen.[9] This Act, passed alongside the Census and Statistics Act for India,[10] recognizes that civil registration and vital statistics (CRVS) systems are a key both to efficient public administration as well as to maintaining the human rights to documented citizenship and social security.[11] Given the logistical challenges associated with registering all births and deaths across India, the Indian government historically developed and maintained a number of overlapping health information systems to meet these needs. In this section, I discuss three information systems that are crucial to understanding child mortality across India: the Sample Registration System, the Civil Registration System, and a system of regular household surveys focused on maternal and child health.

Following the passage of the 1969 Registration of Births and Deaths Act, the Indian government implemented the Sample Registration System (SRS) in 1970, then expanded it to a nationwide system in 1976-77.[12] As of 2017, the SRS covered a population of approximately 7.9 million people (approximately 0.6% of the population of India) across 3,892 urban and 4,961 rural sampling units.[13] The system is based on a two-tiered sampling strategy: when a new sampling unit is added, a baseline census of the area is taken. Vital events are then registered continuously, with a survey of the sampling area conducted every six months for an independent count and demographic update.[14] Indicators of fertility, population, and age-specific mortality are then generated at the national and state levels, disaggregated by urban and rural status. Infant mortality is also reported at the sub-state natural division level for select large states.[13]

Conversely, the Indian Civil Registration System (CRS) aims for universal registration of all births and deaths across India. Complete coverage of the CRS is the pathway by which all Indian citizens can access the legal and civil protections afforded by birth and death registration, as guaranteed by the 1969 Registration of Births and Deaths Act.[9,15] Registration in the CRS relies on individual reporting of vital events, which falls to household heads in cases where births and deaths occur in the home, or to facility heads for vital events that occur in institutional settings. This self-report principle can be challenging given the large number of births and deaths that occur at home or in private facilities.[16] To partially address these challenges, the Indian government has targeted improvements in the coverage of CRS birth an death registration based in part on digital registration; however, the estimated completeness of the CRS still remains low compared to the SRS.[17]

Three major survey series conducted by the Indian government also capture aspects of maternal and child mortality nationwide. These series are the District-Level Household Surveys (DLHS); the National Family Health Surveys (NFHS), conducted in partnership with the international Demographic and Health Surveys program; and the Annual Health Surveys (AHS). Of these, the AHS is the largest, with over 4.3 million households captured in its 2013 sample.[18] While these surveys use differing sampling strategies, survey different household members, and collect information on diverse topics, all three capture retrospective information on child mortality by requesting birth histories from women of reproductive age.[18] While both the DLHS and NFHS surveys are conducted in five-year intervals, a two-year gap between them offers temporal survey coverage that is unparalleled in other countries. However, survey-specific differences in questions related to adult mortality and non-communicable disease burden limit their utility for understanding the course of the epidemiological transition across India.[5]

Among these data sources, the SRS is considered to be the gold standard for estimating fertility, births, and deaths across the country due to its relatively large sample size and representative sampling design.[14] Past studies have used the SRS as the baseline against which the completeness of CRS mortality reporting is estimated.[17] Against this standard, the coverage of the CRS has been increasing: using SRS as the baseline, a study estimated that CRS coverage increased from 55% to 77% nationally, with completeness approaching 100% in nine states.[17] However, the Register General of India, which maintains the SRS, has only infrequently published estimates of the source's completeness. An independent investigation of SRS completeness using the Brass Generalized Growth Balance method found that at the state level, the completeness of death registration in the SRS between 1981-1990 varied between 81% and 100% for males and 74% to 95% for females across all age groups.[12] A more recent study of SRS completeness using the Preston and Coale method estimated that SRS completeness varied between 77% and 99% between 1990 and 2007 without a clear trend showing improved completeness over time. Notably, this study also found that SRS in the southeastern Indian state of Andhra Pradesh captured only 58% of deaths in 2007, the final year of estimation.[14]

As tools for identifying health disparities nationwide, the SRS and CRS are limited in the spatial and sub-population data they report. Neither of these two sources reports mortality at the sub-state level, except for the infant mortality rate (IMR), which the SRS has begun to report by district grouping in recent years.[14] The survey systems report township-level geographic information associated with each sample cluster; however, these surveys capture only retrospective information about child mortality that is subject to possible response

4

biases and do not include information about causes of death that are provided by CRS records.[18]

2 Methods

In this chapter, I develop a space-time modeling technique to estimate spatial variation in child mortality using data from the three major household survey series. This model takes advantage of the precise spatial information offered by the household surveys while attempting to correct for possible recall and omission biases associated with retrospective survey data. I then compare these modeled estimates of child mortality to the SRS and CRS data sources to explore possible cross-source differences in mortality estimates. While I initially prepared this model as part of a collaborative project estimating causes of child mortality across India,[19] I produced all survey-based estimates of child mortality reported in this chapter. The comparison of estimates between SRS and survey-based data presented below are novel to this thesis.

2.1 Geo-locating complete birth history data

The NFHS, DLHS, and AHS all collect complete birth history from women of reproductive age. Complete birth history data provide the month and year of birth and death for each child of an interviewed woman. Complete birth histories from individual-level survey data were extracted between 2000 and 2017, then reshaped these birth histories to reflect the number of children entering distinct age groups between birth and age five, as well as the number who died within each age group, by survey cluster and retrospective year.[20,21] Survey clusters were then matched to precise spatial identifiers such as GPS points, precise township names, or districts. Across the three survey series, 3.3 million georeferenced birth histories were incorporated into the spatial model. Figure 1 shows the distribution of retrospective birth histories by year of birth across the study time period.

2.2 Space-time-age mortality estimation model

To synthesize information across various sources, and to make consistent estimates across space and time, I fitted a discrete hazards geostatistical model to the data. Age groups were represented in seven mutually exclusive bins (0, 1–5, 6–11, 12–23, 24–35, 36–47 and 48–59 months), each with a baseline mortality probability that was assumed to be constant nationwide. The model explicitly accounted for variation across age bin, year and space through inclusion of both fixed and random effects. Indicator variables for each age bin were included to form a discrete baseline mortality hazard function, representing the risk

Sample size for retrospective birth histories over time
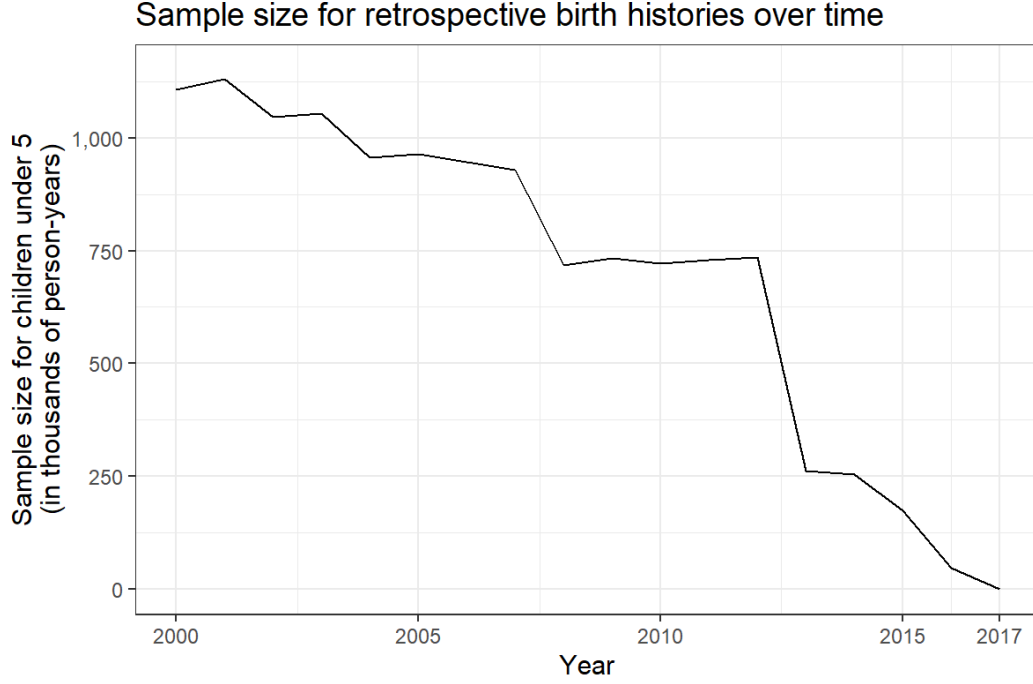
Figure 1: Sample size for under-5 mortality modeling based on survey data between 2000 and 2017 across India. The sample size for each year is approximated here as the number of individuals under age 5 entering each year according to retrospective complete birth history data. At the time of analysis, no complete birth history data was available after 2016, the year of the most recent included household survey.

of mortality in discrete bins from birth to 59 months of age with covariates set at their means. Baseline hazard functions were allowed to vary in space and time in response to changing covariate values, as well as in response to linear effect on year. These estimated fixed effects were then applied to the gridded surface of covariate values to make predictions across the entire study geography. A latent process effect was also included to account for remaining correlation across age, time and physical space after accounting for fixed effects and source-specific biases. As such, estimates at a specific age, time or place benefitted from drawing predictive strength from data points nearby in all of these dimensions.

All data were prepared such that we counted or estimated the number of children entering into (n) and dying within (Y) each period–age bin from each GPS-point location (s) in each survey (k). The number of deaths for children in age band (a) in year (t) at location (s) was assumed to follow a binomial distribution:

$$Y_{a,s,t} \sim Binomial(n_{a,s,t}, P_{a,s,t})$$

6

where $P_{a,s,t}$ is the probability of death in age bin (a), conditional on survival to that age bin for a particular space–time location. Using a generalized linear regression modelling framework, a logit link function is used to relate P to a linear combination of effects:

$$logit(P_{a,s,t}) = \beta_0 + \beta_1(a) + \beta_2 X_{s,t} + \beta_3 t + \nu_k + Z_{a,s,t}$$

The first term, β_0, is a global intercept, representing the mean for the first age band when all covariates are equal to zero, whereas $\beta_1(a)$ are fixed effects for each age band, representing the mean overall hazard deviation for each age band from the intercept, when all other covariates are equal to zero. β_2 are the linear fixed effects of geospatial covariates $(X_{s,t})$, while β_3 is a linear temporal effect by year. The term $\nu_k \sim Normal(0, \sigma_k^2)$ is a survey-level random effect used to account for systematic variation or biases across data sources: this term was fit to the deviation between overlapping survey estimates, and was excluded from the final model predictions as a bias term. Finally, the term $Z_{a,s,t} \sim GP(0, K)$ is a correlated four-dimensional separable Gaussian process, accounting for structured residual correlation across the indices of space, time, and age that are not accounted for by any of the model's other effects. This structure formalizes Tobler's first law of geography and extends it to other dimensions: all observations in space-time-age are related, but observations that neighbor each other in these dimensions are more likely to take similar values than observations that are distant from each other. The covariance matrix K is constructed as a separable process across age, space and time $(K = \Sigma_a \otimes \Sigma_t \otimes \Sigma_s)$. The continuous spatial component is modelled with a Matérn covariance function, and the age and temporal effects were each assumed to be discrete auto-regressive order 1. This model draws from previous studies investigating variation in child mortality based on survey data, particular Burstein et al.'s study in low- and middle-income countries.[20,22]

2.3 Mortality forecasting

After estimating NMR, IMR, and U5MR between 2000 and 2017, I calculated a weighted annualized rate of change for all grid cells and draws, giving greater weight to more recent annual rates of change. I then applied this annual rate to each indicator cell-draw in the final year of mortality estimates, in 2017, to project estimates for all three indicators through 2025 while preserving uncertainty. Previous spatial analyses have established this method for projecting spatial estimates into the future.[23] I then applied a population-weighted aggregation to calculate estimated mortality at the district and state levels based on gridded mortality projections.

2.4   Comparisons to SRS data at the most detailed spatial level available

In recent years, the Indian Sample Registration System has begun to report estimates for
infant mortality rates across 68 sub-state "natural divisions" in the larger states of India.[13] I
used population-weighted aggregation to estimate infant mortality by natural division, with
uncertainty, in 2017. I then compared these estimates based on survey data with point
estimates provided by the SRS at the same spatial resolution in 2017. Rather than using
either source as a "gold standard", I review significant differences between survey and SRS
data as a starting point for future research and improvement.

3   Results

A small-area analysis of mortality reveals focal areas and within-state disparities in child
mortality that would be obscured in a state-level analysis. While the neonatal, infant, and
under-5 mortality rates dropped in almost all districts of India between 2000 and 2017, stark
inequalities in child survival remain across the country. Simple projections to 2025 reveal
that while absolute differences in mortality across district will generally converge in future
years if current trends persist, many districts are not on track to meet the 2025 child survival
goals set by the National Health Policy.

A comparison of infant mortality estimates between the Sample Registration System and
synthesized survey data by Indian natural division reveals that both sources identify the
same focal regions for infant mortality nationwide. The Sample Registration System es-
timates significantly lower infant mortality rates than survey-based estimates in 11 of 68
natural divisions, primarily in states where past analyses have identified deficiencies in SRS
completeness. However, the larger sample size of the SRS facilitates precise estimates of in-
fant mortality in each natural division, complementing a weakness of survey-based estimates
at the end of the study time period.

3.1   Progress and local disparities in child survival

A synthesis of survey data indicates that at the national level, the under-5 mortality rate
across India fell to 42.4 (37.2-49.0) per 1,000 live births by 2017, from 83.1 (77.6-88.6) in
2000. Across all states in the country, child mortality varies almost 6-fold, from 10.4 (7.5-
14.2) in Kerala to 59.7 (50.4-71.8) in Uttar Pradesh. This inequality, while stark, pales in
comparison to differences at the district level. Figure 2 shows the distribution of under-5
mortality across India, which ranged from a low of 16.4 (13.2-20.2, in Thrissur district in
Kerala) to a high of 163.3 (147.4-179.6, in Panna district of Madhya Pradesh) in 2000 and

from 8.4 (5.6-12.1, again in Thrissur district) to 87.9 (72.2-106.4, in Budaun district in Uttar Pradesh) in 2017. In 2017, under-5 mortality in 86 of the 723 districts nationwide exceeded the highest state-level U5MR; of these, more than half were found outside of Uttar Pradesh state, including 14 districts in Assam and 14 in Madhya Pradesh.

In many ways, estimates of neonatal and infant mortality rates from 2000-2017, also shown in Figure 2, reflect a similar pattern of unequal progress. Between 2000 and 2017, the neonatal mortality rate declined nationally from 37.8 (35.2-40.3) to 23.3 (20.4-26.9), while the infant mortality rate declined from 62.7 (58.5-66.9) to 36.0 (31.6-41.5). However, at the district level, neonatal mortality ranged 8-fold, from 5.8 (3.9-8.2) in Ernakulum, Kerala state to 46.2 (37.5-56.6) in Budaun, Uttar Pradesh. District-level rates of infant mortality ranged 10-fold, from 7.3 (4.9-10.6) in Thrissur district, Kerala state to 75.9 (62.1-92.3) in Budaun, Uttar Pradesh. As with under-5 mortality, districts with the highest rates of infant mortality were generally concentrated in the states of Uttar Pradesh, Madhya Pradesh, Chhattisgarh, and Rajasthan in 2017.

While the under-5 mortality rate declined by 49 percent from 2000 to 2017, equivalent to a 3.8% annualized rate of decline, the speed of improvements in child survival varied considerably across the country. Figure 3 shows the annualized rate of decline in child mortality by district from 2000 to 2017. Across 103 districts concentrated along the southeastern coast, Telangana state, and Arunachal Pradesh, annual decline in child mortality exceeded 5% per year, equivalent to a nearly 60% reduction in U5M across the study period; meanwhile, 23 districts concentrated primarily in Rajasthan and Himachal Pradesh in the northwest and Mizoram and eastern Assam in the northeast experienced less than 2% of annual decline in the child mortality rate. When observing the pattern of mortality decline across Indian districts, the space-time pattern in the northeast of the country is of particular interest: although Arunachal Pradesh and western Assam have experienced rapid declines in mortality, child mortality remains stubbornly high in a region covering eastern Assam, Mizoram, and Tripura. A number of factors could be influencing this unusual time trend across Northeastern India: the region's geographical separation from the rest of the country, heterogeneous ethnic makeup, proximity to Myanmar and Bangladesh, and high prevalence of drug-resistant malaria could all play a role in shaping under-5 mortality.[24,25] Further investigation is needed to understand why gains in child survival observed in Arunachal Pradesh and western Assam have not been matched in other Northeast Indian districts.

The differential improvement in child survival across the eastern and western halves of Assam is one of many insights revealed by a district-level analysis. Figure 2 also demonstrates spatial trends in neonatal, infant, and child mortality that would be obscured by a state-

Neonatal (<1 month)
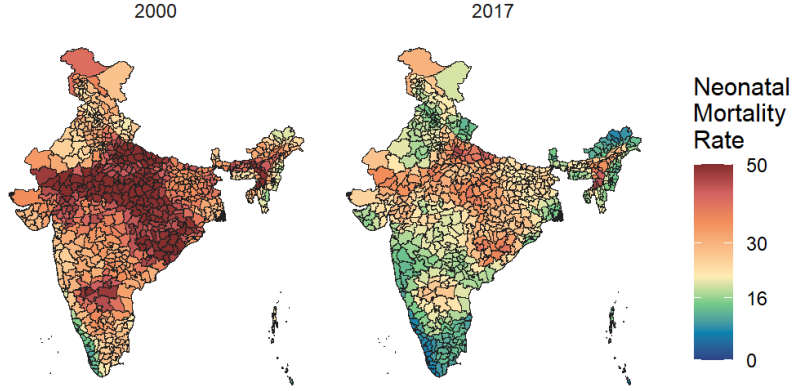
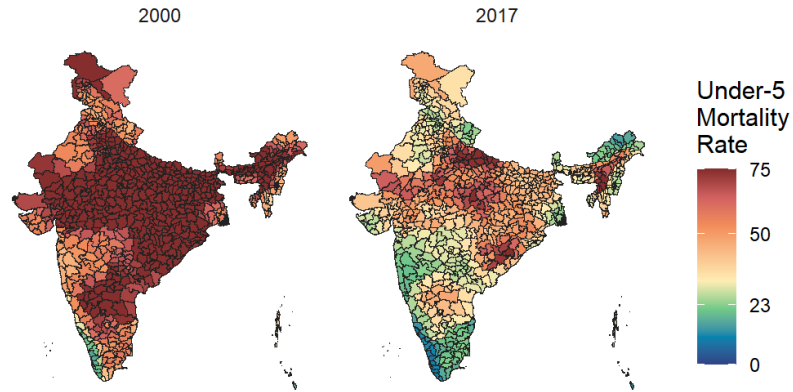
Infant (<1 year)
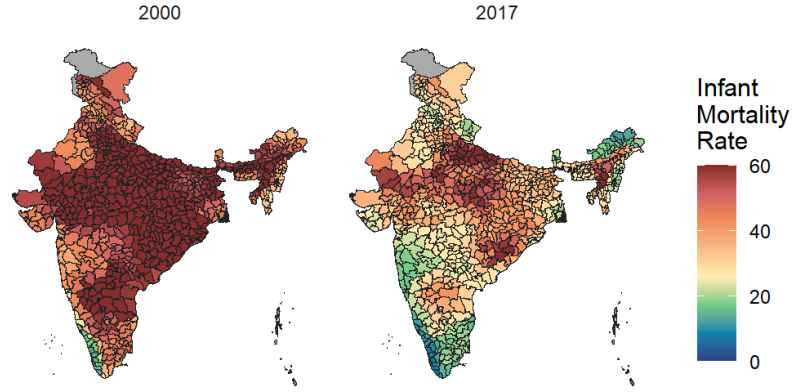
Child (<5 years)

Figure 2: NMR, IMR, and U5MR per 1,000 live births in 2000 and 2017, estimated from survey data. The Indian National Health Plan 2017 calls for reducing neonatal mortality to less than 16 deaths per 1,000 live births, and under-5 mortality to less than 23 deaths per 1,000 live births, nationwide by 2025. The color scale for mortality varies across age groups, with green centered around the NHP mortality targets for each age group.
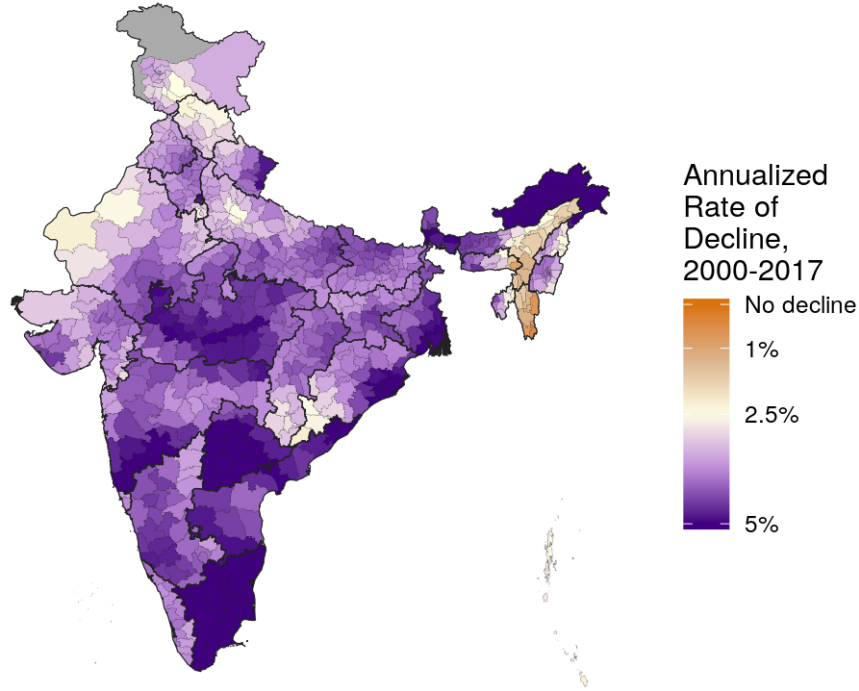
Figure 3: Annualized rate of decline (ARoD) in U5MR between 2000 and 2017. An ARoD of 5% is equivalent to a cumulative decline of 58% over 17 years, while an ARoD of 1% is equivalent to a cumulative decline of 16% over 17 years. ARoD rankings by district were stable across age-specific indicators.

level analysis. The states of Karnataka, Maharashtra, and Gujarat experience significantly lower NMR, IMR, and U5MR in the coastal east than in the inland west. In the state of Rajasthan, U5MR is approaching the NHP target of 23 in 6 districts near Jaipur city, while 10 districts in the southwest of the state still suffer from a U5MR exceeding 60. The ratio between the under-5 mortality rates of any two districts in the same state reaches 3.6 (in Karnataka state, from 11.3 (7.8-15.7) in Dakshin Kannad district to 41.2 (30.4-54.2) in Koppal district). A similar ratio of 3.7 is observed across neonatal mortality rates in the districts of Karnataka, suggesting that many of the inequalities observed in child survival arise from disparities in health burden within the first month of life.

This space-time model also offers insight into changing within-state inequalities over time. Figure 4 shows the spread of NMR, IMR, and U5MR across districts within each Indian state and union territory in 2000 (grey) and 2017 (blue). Each dot represents a district: the lower bound of each vertical spread represents the district with the lowest U5MR in each state and year, while the upper bound of each vertical spread represents the district with the highest U5MR in the same district and year. The difference in the U5MR between

11

the highest and lowest districts can be interpreted as one measure of absolute inequality in U5M within each state: by this metric, absolute inequality in U5M declined in almost all Indian states between 2000 and 2017. Figure 4 reveals the diverse relationship between declining state-level U5MR and increasing equity over the study time period. For example, in Uttar Pradesh, state-level U5M declined by 47%, from 112.8 (105-120.5) to 59.7 (50.4-71.8), but absolute inequality only declined slightly, from a difference of 63 to 46. In Bihar, state-level U5MR and inequality declined considerably between 2000 and 2017: state-level U5MR dropped 51%, from 88.8 (82-95.6) to 43.8 (35.9-54), while the U5MR spread across all districts declined from 31 to 13. Several northeastern states were notable exceptions to the trend of declining inequality: Meghalaya and Assam both displayed greater between-district inequalities in 2017 than in 2000.

Although the NMR, IMR, and U5MR all represent probabilities of death rather than true rates, the number of deaths in each age group can be estimated by employing a simple demographic transformation and multiplying by underlying child populations in each district derived from WorldPop.[20,26] Estimates of the number of neonatal, infant, and under-5 deaths by year are therefore functions of mortality, district size, and population density in each district. Nevertheless, estimates of death counts are useful guides for child health interventions that can only cover a limited geographic area. In 2017, an estimated 1.03 million (.90-1.20 million) children under the age of 5 died across India. The three districts with the highest number of deaths were Allahabad (10,800 (8,300-13,800) deaths), Sitapur (8,600 (6,900-10,800) deaths), and Bareilly (8,400 (6,800-10,400) deaths), all in Uttar Pradesh: this largely reflects the large child populations of these three districts as well as their exceptionally high child mortality rates. Counterfactual analyses of child mortality rates can also be used to estimate possible lives saved under different interventions. For example, if all Indian districts were to meet the NHP 2025 goal of 23 or fewer child deaths per 1,000 live births, without any change in districts that have already met this goal, the lives of 480,000 children could be saved annually.

3.2 Comparison to 2025 and 2030 mortality targets

The left side of Figure 5 shows the mean projected under-5 mortality rate for each district in 2025. If under-5 mortality continues to decline at the same rate as it did in most districts from 2000-2017, with emphasis on the trend of decline in recent years, by 2025 only 44 of 723 districts are estimated to have a U5MR greater than 50 deaths per 1,000 live births. The coastal regions of the country as well as Andhra Pradesh are projected to have the lowest under-5 mortality rates in the country, with 22 districts having a mean projected U5MR of
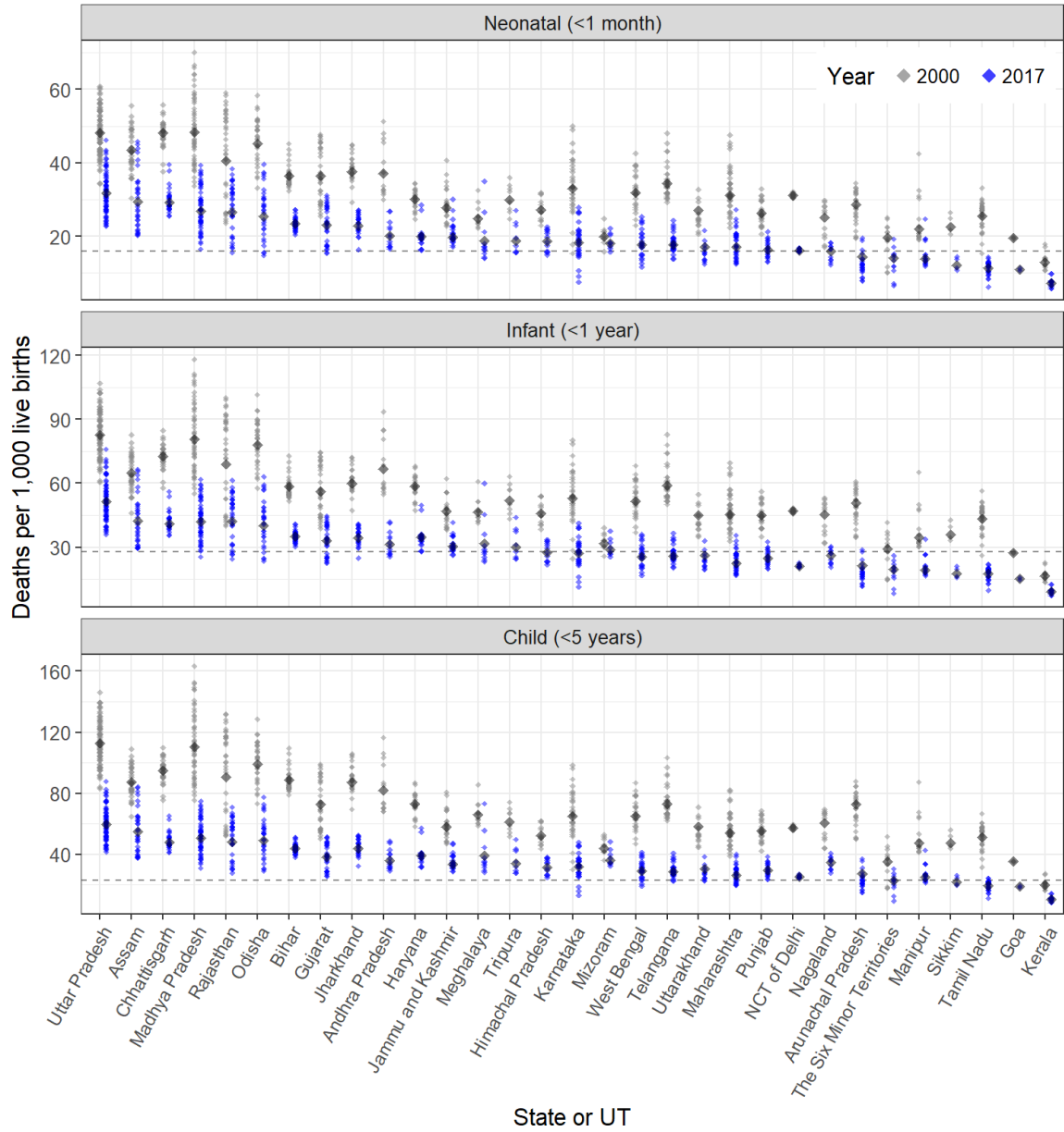
Figure 4: Absolute inequalities in U5MR, IMR, and NMR across districts within each Indian state and union territory in 2000 (grey) and 2017 (blue). Each dot represents a district: the lower bound of each vertical spread represents the district with the lowest U5MR in each state and year, while the upper bound of each vertical spread represents the district with the highest U5MR in the same district and year. The large diamond in each vertical spread shows the overall mortality rates across the state as a whole. A blue bar that is shorter than its grey counterpart indicates that between-district inequality has narrowed between 2000 and 2017. The horizontal dashed lines identify NHP targets for each indicator.

less than 10. While in-state inequalities are projected to remain in the central and western states of Rajasthan, Madhya Pradesh, and Uttar Pradesh, absolute differences between the

13

highest-mortality and lowest-mortality districts in these states are all projected to decline between 2017 and 2025. Due to relatively slow rates of mortality decline between 2000 and 2017, eastern Assam and northern Uttar Pradesh are both projected to become the regions with the highest child mortality by 2025.

325 The right side of Figure 5 highlights areas that are likely to meet or not meet the Indian National Health Policy's 2025 target of fewer than 23 deaths per 1,000 live births, accounting for uncertainty in the projections. Districts with a 95% uncertainty intervals (UI) falling exclusively above 23 were marked as unlikely to meet the NHP 2025 goal, while districts with projected UIs falling exclusively below 23 were marked as likely to meet the goal. All 330 other districts were marked in white, indicating that these districts had projected 95% UIs overlapping with the target and could not be assigned to either meeting or missing the target with high certainty. Of India's 723 districts, only 95 were projected to meet the target and 232 were projected to miss the target with high confidence. Only the states of Kerala and Goa and the Union Territory of Puducherry were projected to meet the target 335 in all districts with high confidence. On the other hand, more than half of all districts in Rajasthan, Madhya Pradesh, Uttar Pradesh, Chhattisgarh, Jharkhand, and Odisha were projected to miss the NHP 2025 target with high confidence. Notably, no states included both districts projected to meet and districts projected to miss the target.

Figure 6, below, plots districts ranked by their levels of under-5 mortality in 2000, 2017, 340 and 2025 projections across the states of Andhra Pradesh, Rajasthan, and Telangana. This figure demonstrates that the trend of converging mortality rates across districts, noted above in reference to Figure 4, are projected to continue to 2025 within most states. Uncertainty in district estimates is represented by colored vertical bars. Due to a larger data sample available in the year 2000 compared to 2017 as well as the uncertainty inherent in the 345 projection method, mortality uncertainty intervals tend to widen from 2000 to 2017 and from 2017 to the 2025 mortality projections.

3.3   Comparing local estimates from CRS, SRS, and survey-based estimates

Infant mortality estimates from the Sample Registration System are published at the natural division level across 19 large Indian states. Although smaller than states, natural divisions 350 typically contain several districts: according to WorldPop estimates, the 723 districts of India had an average infant population of 32,000 in 2017, while the SRS reported IMR estimates across 68 natural divisions with an average infant population of 324,000.[26]

In 2017, the SRS estimates the infant mortality rate across India to be 33 per thousand live births, compared to a national IMR of of 36.0 (31.6-41.5) estimated by the survey-based

Under-5 Mortality Rate, 2025 (projected)
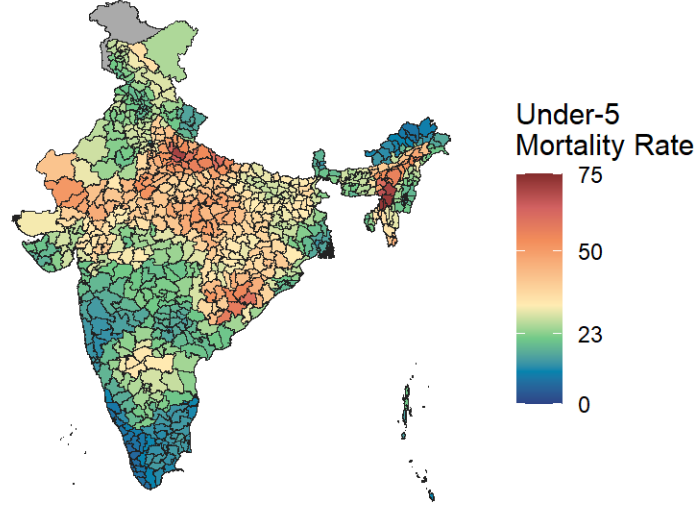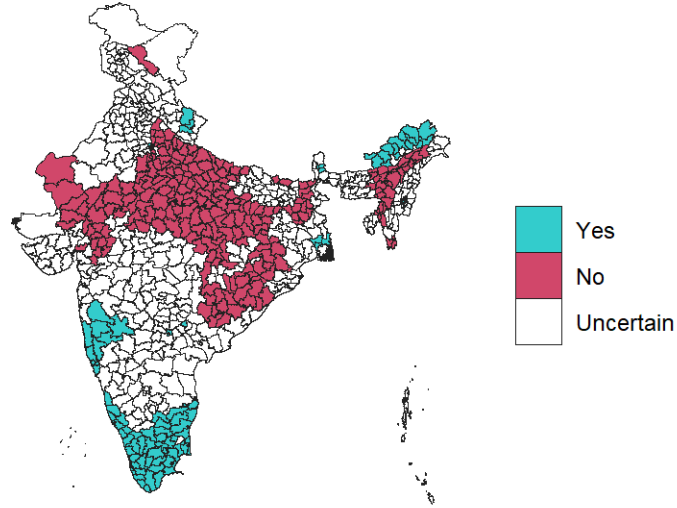

Projected to meet NHP 2025 target?

Figure 5: *Left*: Mean estimated U5MR by district in 2025, projected forward by extending the annualized rate of decline from 2000-2017. *Right*: Comparison between projected 2025 U5MR estimates and the Indian National Health Policy 2017 target of 23 under-5 deaths per 1,000 live births, accounting for uncertainty. Districts shaded in white have a 95% uncertainty interval (UI) that overlaps with the NHP 2025 target; districts shaded in red have a non-overlapping UI that falls above the target; districts shaded in blue have a non-overlapping UI that falls below the target.
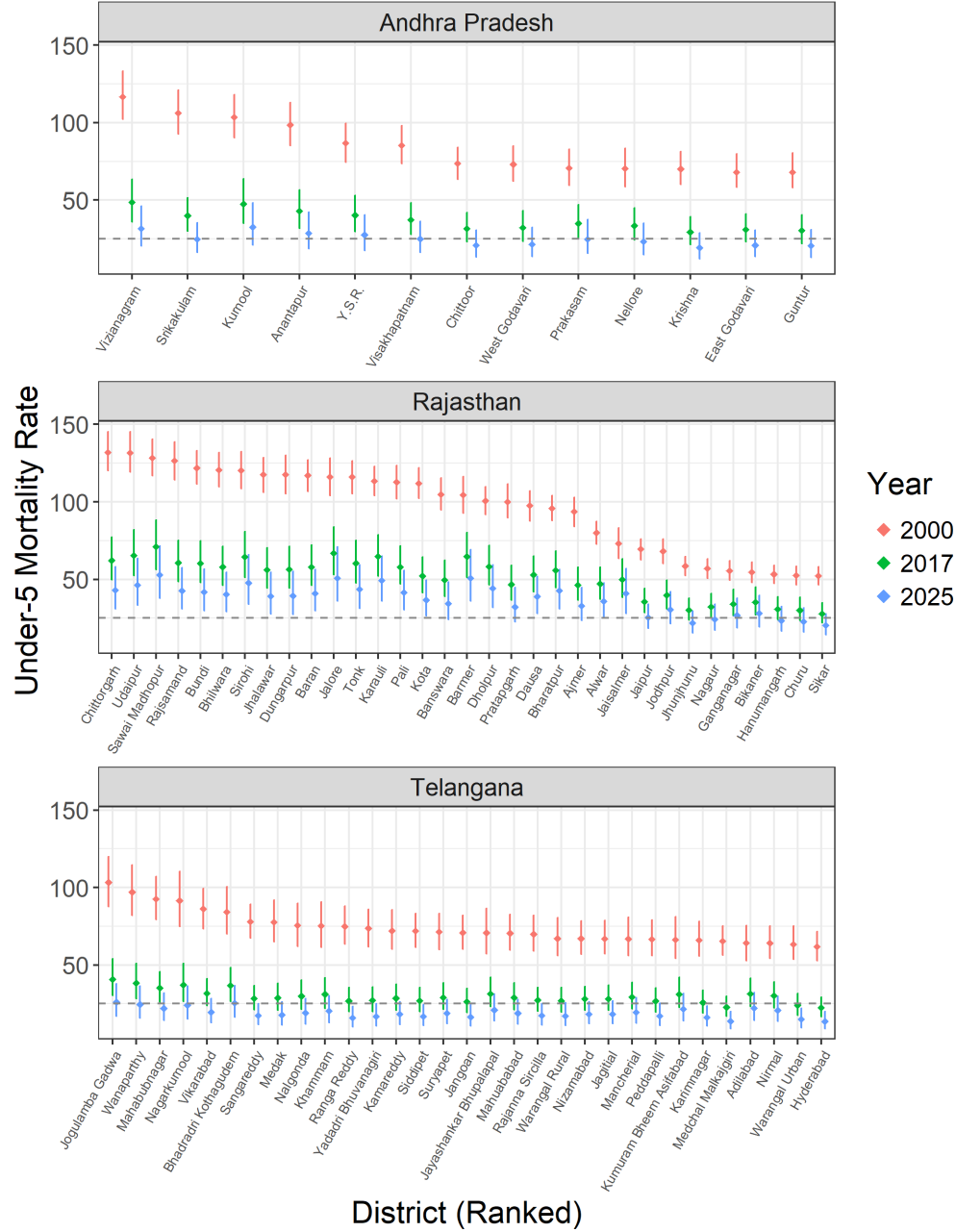
Figure 6: Estimated U5MR by district in 2000, 2017, and projected to 2025 across three exemplar Indian states. Each vertical line displays the bounds of the 95% uncertainty interval in a given district and year, and the point within the vertical line displays the corresponding mean U5MR estimate for that district and year. Districts are sorted in decreasing order of their estimates U5MR in 2000. The dotted horizontal line displays the Indian NHP target of fewer than 23 under-5 deaths per 1,000 live births by 2025.

355   model.[13] Figure 7 Panel A shows variation in infant mortality across India according to SRS data. SRS data shows similar national trends as the survey-based model, with the highest-

mortality areas in the Northern division of Madhya Pradesh state as well as the Central Brahamputra Plains and Cachar Plains divisions in Assam state to the northeast. SRS data shows 7.5-fold variation in infant mortality by division: the lowest-mortality division is the Southern division of Kerala. Although natural divisions obscure local variation occurring at finer spatial resolutions, within-state variation is still apparent in Karnataka state, where recorded IMR varies 2-fold across its constituent natural divisions.

Although the difference in the national IMR calculated between SRS data and the survey-based model falls within the bounds of model uncertainty, a subnational analysis demonstrates why the SRS model estimate falls slightly lower than the mean survey-based estimate. Figure 7 Panel B shows the ratio between the IMR mean estimates derived from SRS versus survey data: from this map, it is apparent that the SRS estimates lower infant mortality rates in almost all natural divisions, with the notable exceptions of Madhya Pradesh and two natural divisions in Assam. In the Northern Upper Ganga Plain division of Uttar Pradesh state and Ladakh, the IMR estimated by the SRS is more than 40% lower than the estimate derived from survey data. Figure 7 Panel C highlights the natural divisions where SRS infant mortality estimates fall outside the 95% uncertainty intervals of survey-based estimates. Of the 68 natural divisions reported by the SRS, only three divisions are significantly higher in SRS than in the survey-based estimated, including the Plains Western division in Assam; but 11 are significantly lower, including the four northernmost natural divisions in Jammu and Kashmir/Ladakh as well as the Western Plains division of West Bengal.

4   Discussion

To develop programs that foster child health across a populous and diverse country, Indian policymakers need health information systems that can identify local variation in child mortality. Estimates of neonatal, infant, and under-5 mortality across the 723 districts of India reveal key trends and focal areas that should inform the implementation of the National Health Policy's child mortality priorities. Given that the probability of a child dying before their fifth birthday ranges from 9 to 88 per thousand (1 to 9 percent) depending on their district of residence, child health program administrators should consider expanding programs in areas where child health burden is highest. This observation also holds for the 36 states and union territories of India, where the under-5 mortality rate can varies up to 6-fold. Only three states and union territories, Goa, Kerala, and Puducherry, are on track to meet the NHP U5MR target of 23 deaths per 1,000 live births in all of their districts with high certainty, meaning there is an urgent need to increase both maternal and child health services as well as local health data collection across most of the country. This analysis also
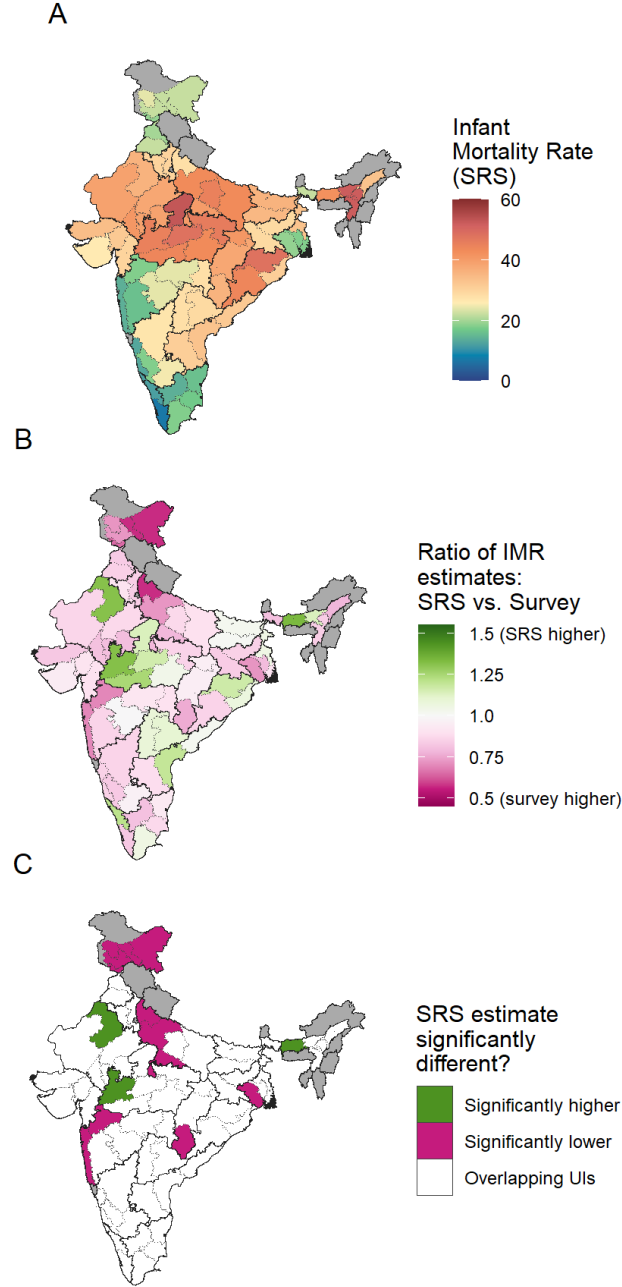
17

A

B

C

Figure 7: *A* : IMR per 1,000 live births in 2017 across 68 natural division in large Indian states, as estimated by the Indian SRS 2017 report. *B* : By aggregating survey-based estimates of IMR to the natural division level, estimates of IMR from SRS and survey sources can be compared. The ratio between SRS-based IMR estimates and survey-based mean IMR estimates are shown at the natural division level. *C* : Comparison between point estimates of IMR provided by the SRS with the 95% uncertainty interval for IMR estimated by survey data. Compared to survey-based estimates in 2017, the SRS estimated significantly lower IMR for 11 of 68 natural divisions, and significantly higher IMR for 3 of 68 natural divisions.

reveals regional hot spots of high under-5 mortality—notably in eastern Assam, Tripura, and Mizoram—that are obscured by state-level estimation projects. These hot spots offer an opportunity for further investigation of the underlying conditions that are impeding progress in child survival.

These insights are enabled by georeferenced birth histories provided by surveyed mothers. These birth histories have several advantage over data from the Sample Registration System and Civil Registration System, particularly the advantage of a known, and complete, denominator for child mortality estimation. However, if reported at the district level, SRS and CRS data could also supplement some of the limitations evident in survey-based estimation. Unlike SRS and CRS data, retrospective survey data suffers from diminished sample sizes in more recent years of reporting, leading to increased uncertainty in the years of analysis with the greatest policy relevance. Additionally, surveys only reliably capture mortality in young age groups, while CRS and SRS data could report age-specific mortality across the entire life course.

4.1 Scale and inequality in mortality

This analysis demonstrates that spatial scale matters for identifying disparities in child survival across India. According to the spatial model presented here, infant mortality varies 5.6-fold at the state level (the level reported by the CRS); 7.3-fold at the natural division (the level reported by the SRS); and 10.3-fold at the district level. As results are presented at more aggregate levels, focal areas with persistently high mortality as well as exemplar areas displaying substantial improvements over time are convolved with other districts, obscuring patterns that could be used to uncover mortality drivers or develop interventions. Given that approximately 160,000 children under 5 live in the average Indian district,[26] this analysis presumably obscures even more local disparities in child mortality.

This analysis also allows policymakers to target and prioritize interventions to specific districts with the highest mortality rates or greatest number of child deaths. As shown above, both mortality rates and counts are distributed highly unequally across the country and within states. Given previous evidence about the context dependence of child health burden across India, interventions must be tailored to survival in particular age groups and urban-rural context, balancing between the priorities of efficacy and equity in service delivery.[3,6]

4.2 Limitations of survey-based spatial mortality mapping

As the only data source that includes the spatial identifiers needed to conduct a district-level analysis, retrospective birth histories from India's three household survey series were

essential to this analysis. In addition to their spatial metadata, these surveys facilitate
analysis thanks to a set of questionnaires designed to minimize bias, reporting of sampling
weights associated with each household, and a documented methodology that is comparable
across survey series and even across countries.[18,27] However, survey-based spatial models of
mortality are also subject to limitations inherent in the data and modeling approach that
limit some aspects of their policy utility.These limitations could be alleviated if SRS or CRS
data was made available to compare with survey-based estimates.

The first programmatic challenge of this survey-based model relates to data availability
over time. Although policy makers tend to be most interested in the most recent year of
analysis, birth history, as a retrospective data source, increases in quantity farther back in
time. Figure 1, which shows the combined sample sizes of all birth history survey sources
used in this analysis, demonstrates this issue across India between 2000 and 2017. Thanks
to retrospective data collection, birth histories covering the year 2000 are available from all
eight DLHS, NFHS, and AHS surveys collected since the year 2000. Conversely, in 2016,
only a partial year of data is available from the 2016 NHFS—and, in the final year of analysis,
no retrospective data is available. While the spatial model leverages known relationships
over the dimensions of space, time, and age and with covariates to produce estimates of
mortality in the final year, this estimate tends to be more uncertain due to the scarcity of
underlying data.

In many Indian states, decreasing sample sizes and converging between-district mortality
rates complicate district rankings in 2017 as well as projections forward to 2025. Figure
6 demonstrates this limitation in three states by showing under-5 mortality estimates by
district in 2000, 2017, and 2025 for three states. Districts are ordered according to their
estimated under-5 mortality rate in 2000. The states of Andhra Pradesh, Rajasthan, and
Telangana, showed stark disparities in U5MR between their lowest- and highest-burden
districts in 2000. These disparities, while reflecting an unconscionable situation on the
ground, facilitate making clear and highly-certain distinctions between areas with low and
high mortality within a state: for example, the difference in U5MR between the districts of
Chittorgarh and Sikar in Rajasthan was 80 per thousand in the year 2000. The progress
that all three states have made in reducing between-district mortality inequality, while
laudible, increases the analytical difficulty of identifying the highest-mortality districts in
each state in 2017. Ranking the relative mortality of districts in 2025 becomes even more
difficult given both the convergence in mortality rates and uncertainty in the underlying
annualized rates of change from 2000 to 2017, a symptom of a shrinking sample size. While
current and future district rankings are a highly sought-after metric for targeting program

funding within states, these rankings should be communicated extremely carefully due to their inherent uncertainty.

Another concern relates to potential bias in unsampled locations, which can be sensitive to model specification. The spatial model described in this chapter makes predictions across 5 x 5 km grid cells for each estimated year and age group. In location-age-years not overlapping sampled data points, the model estimates the underlying mortality level using covariate information and a latent surface that is influenced by nearby observations in the dimensions of space, time, and age.[20,28] This introduces concerns that relationships between child mortality and underlying covariates may not hold outside of the sampled data points. In the current model, covariates are capped at their minimum and maximum observed values: this avoids unfounded extrapolation of covariates at unobserved values, but may also miss relationships at the extremes of child mortality that were not found in any sampled location. While different functional forms can be applied to estimate the relationship between mortality and underlying covariates, additional model flexibility can come at the expense of interpretability in a policy context.[29,30] SRS and CRS data sources can alleviate the issue of extrapolation: because every district is surveyed, they open up the possibility of using small-area model frameworks where data has been sampled across every space-time-age combination in the study area.[31]

4.3   Opportunities offered by spatial reporting of CRS and SRS data

When SRS and a survey-based model produce conflicting estimates of infant and child mortality across India, it begs the question: which source is correct? Other studies comparing multiple sources of mortality data across India have measured data quality against a selected "gold standard."[17] When comparing SRS and survey-based estimates, I assert that survey data should be used as a gold standard, subject to the limitations of model specification and uncertainty. While the Sample Registration System has a larger sample size than all household surveys combined, and includes a census component that is intended to ensure completeness, past evidence as recently as 2007 suggests that in practice, past rounds of SRS data can be highly incomplete.[12,14] Mortality data in the SRS is also sensitive to the population denominator in each sampled location, which may miss early neonatal deaths as well as migration in and out of the study area. While survey-based estimates are limited by relatively small sample sizes in recent years, this disadvantage should be reflected in the uncertainty surrounding mortality estimates in 2017, given that the model is correctly specified. This chapter identified 14 natural divisions where SRS estimated infant mortality rates falling outside of the 95% uncertainty intervals estimated by the survey-based model:

these differences offer an opportunity for further investigation, with an emphasis on barriers that may be impeding complete reporting for the SRS.

495 Despite these limitations, the SRS is still the most promising data source for future district-level analysis of mortality across all age groups in India. Birth history data is generally used only to predict mortality rates among children, due to sparse data and potential recall bias for predicting mortality among older age groups.[22] Conversely, SRS data completeness is likely to be higher for reported adult deaths than among children. Lower completeness in 500 under-5 mortality reporting may be due in part to different cultural conceptions of mortality in the neonatal context or social insurance programs offered to the relatives of deceased adults.[17] If SRS data was reported at the district level or by sampling unit, its much larger sample size and annual time trends would enable policymakers to assess spatial variation in mortality with greater precision, and to more confidently assess progress towards future mor-505 tality targets. It is a near-certainty that adult mortality, like child mortality, is distributed unequally across the country in patterns not fully reflected by state-level tabulations. Even an imperfect accounting of these differences would greatly increase our understanding of the diverse epidemiological contexts across India.

When discussing estimates of all-cause mortality rates, the discussion so far has centered 510 around survey data and the Sample Registration System. When estimating all-cause mortality rates, the Indian Civil Registration System is widely understood to be incomplete compared to these other two data sources.[17,32] To understand why the CRS is a vital component needed to achieve the National Health Policy, we must turn to the four principles of a successful vital registration system: completeness, accuracy, universality, and confidentiality.[33] 515 While source-specific completeness has been one focus of this chapter, detailed CRS tabulations could also inform mortality estimates by acting as a floor for the total number of deaths that occurred in a given district, age group, and year. However, we should understand the CRS not just as a source of health statistics, but as the foundation for legal rights offered to all Indian citizens that can only be obtained through birth and death registration. When 520 it comes to the principle of universality, and the legal right of all Indians to access the benefits of civil registration, only the CRS is designed to ultimately cover all residents of India. The CRS also adds cause-of-death assignment as another dimension of mortality reporting. While cause-of-death classification accuracy remains challenging for deaths that occurred in the home,[34] even flawed data on causes of mortality provides a wealth of information 525 about local variation in the epidemiological transition taking place across the country. Finally, while local reporting of SRS and CRS data would enable new policy-relevant insights about the state of health across India, strict practices must be maintained to ensure that

22

all records remain confidential. Reporting standards for spatially-resolved SRS and CRS data could follow restrictions developed in other countries, such as the United States, that require aggregation of any mortality data where counts fall below a fixed threshold.[35]

4.4   Conclusions

India's adoption of the 2017 National Health Policy indicates the need for local estimates of health burden, particularly among children in the first five years of life. Despite this need, no data system currently provides district-level estimates of neonatal, infant, and child mortality. In this chapter, I develop a model for district-level mortality estimation based on household surveys, the only data source available at the district level and below. This model provides new insights that can inform the drive to meet neonatal and under-5 mortality targets by 2025; however, spatially-resolved data from the Sample Registration System is better-suited to describe local variation in mortality among older age groups. Ultimately, only the birth and death certificates administered by the Civil Registration System can extend the full legal rights afforded by vital records to all Indian citizens.

This chapter has also investigated the relationship between household survey data and the Sample Registration System across 68 natural divisions of India. Comparisons between the two data sources were limited by the relatively large size of each natural division, which concealed local spatial variation; instead, this analysis relied on past literature estimating the relative quality of SRS data. In the following chapter, I demonstrate how mortality data from vital statistics and household surveys can be combined to generate more robust estimates of all-cause mortality.

5   References

1.   Government of India Ministry of Health and Family Welfare. National Health Policy 2017. 28 https://www.nhp.gov.in/nhpfiles/national_health_policy_2017.pdf (2017).

2.   Sundararaman, T. National Health Policy 2017: a cautious welcome. Indian Journal of Medical Ethics 2, 69–71 (2017).

3.   Government of India Ministry of Health and Family Welfare. Situation Analyses: Backdrop to the National Health Policy 2017. 14 (2017).

4. Dicker, D. et al. Global, regional, and national age-sex-specific mortality and life expectancy, 1950–2017: a systematic analysis for the Global Burden of Disease Study 2017. The Lancet 392, 1684–1735 (2018).

5. Yadav, S. & Arokiasamy, P. Understanding epidemiological transition in India. Global Health Action 7, (2014).

560 6. Dandona, L. et al. Nations within a nation: variations in epidemiological transition across the states of India, 1990–2016 in the Global Burden of Disease Study. The Lancet 390, 2437–2460 (2017).

7. World Health Organization. Monitoring the building blocks of health systems: a handbook of indicators and their measurement strategies. vol. 35 1–92 http://www.annualreviews.org/doi/10.1146/annurev.ecolsys.35.021103.105711 (2010).

8. Abouzahr, C. & Boerma, T. Health information systems: the foundations of public
565   health. Bulletin of the World Health Organization 83, 578–583 (2005).

9. Parliament of the Republic of India. The Registration of Births and Deaths Act, 1969. vols Act 18 (1969).

10. Subramanian, S. A Census and Statistics Act for India. Calcutta Statistical Association Bulletin 18, 85–96 (1969).

570 11. UN General Assembly. Universal declaration of human rights. vol. 2 (1948).

12. Bhat, P. N. M. Completeness of India's sample registration system: An assessment using the general growth balance method. Population Studies 56, 119–134 (2002).

13. Office of the Registrar General & Census Commissioner. Sample Registration System Statistical Report 2017. 337 http://www.censusindia.gov.in/vital_statistics/SR
575   S_Report/9Chap%202%20-%202011.pdf (2017).

14. Mahapatra, P. An Overview of the Sample Registration System in India. in Prince mahidol award conference & global health information forum 1–13 (Institute of Health Systems, 2010).

15. Abouzahr, C. et al. The way forward. Lancet Health Metrics Network, World Health Organization 370, 1791–99 (2007).

16. Mohanty, I. & Gebremedhin, T. A. Maternal autonomy and birth registration in India: Who gets counted? PLoS ONE 13, 1–19 (2018).

17. Kumar, G. A., Dandona, L. & Dandona, R. Completeness of death registration in the Civil Registration System, India (2005 to 2015). Indian Journal of Medical Research 149, 740 (2019).

18. Dandona, R., Pandey, A. & Dandona, L. A review of national health surveys in India. Bulletin of the World Health Organization 94, 286–296A (2016).

19. Dandona, R. et al. Subnational mapping of under-5 and neonatal mortality trends in India: the Global Burden of Disease Study 2000–17. The Lancet 395, 1640–1658 (2020).

20. Burstein, R. et al. Mapping 123 million neonatal, infant and child deaths between 2000 and 2017. Nature 574, 353–358 (2019).

21. Ahmad, O. B., Lopez, A. D. & Inoue, M. The decline in child mortality: a reappraisal. Bulletin of the World Health Organization 78, 1175–91 (2000).

22. Wakefield, J. et al. Estimating under-five mortality in space and time in a developing world context. Statistical Methods in Medical Research 28, 2614–2634 (2019).

23. Osgood-Zimmerman, A. et al. Mapping child growth failure in Africa between 2000 and 2015. Nature 555, 41–47 (2018).

24. Ghosh, R. Child mortality in India: A complex situation. World Journal of Pediatrics 8, 11–18 (2012).

25. Zomuanpuii, R. et al. Epidemiology of malaria and chloroquine resistance in Mizoram, northeastern India, a malaria-endemic region bordering Myanmar. Malaria Journal 19, 1–11 (2020).

26. Tatem, A. J. WorldPop, open data for spatial demography. Scientific Data 4, 2–5 (2017).

27. Corsi, D. J., Neuman, M., Finlay, J. E. & Subramanian, S. V. Demographic and health surveys: A profile. International Journal of Epidemiology 41, 1602–1613 (2012).

28. Diggle, P. J. & Giorgi, E. Model-based geostatistics for prevalence mapping in low-resource settings. Journal of the American Statistical Association 111, 1096–1120 (2016).

29. Pletcher, S. D. Model fitting and hypothesis testing for age-specific mortality data. Journal of Evolutionary Biology 12, 430–439 (1999).

30. Lucas, T. C. D. A translucent box: interpretable machine learning in ecology. Ecological Monographs 90, 1–55 (2020).

31. Wakefield, J., Okonek, T. & Pedersen, J. Small Area Estimation for Disease Prevalence Mapping. International Statistical Review insr.12400 (2020) doi:10.1111/insr.12400.

32. Gupta, M., Rao, C., Lakshmi, P., Prinja, S. & Kumar, R. Estimating mortality using data from civil registration: a cross-sectional study in India. Bulletin of the World Health Organization 94, 10–21 (2016).

33. United Nations Statistics Division. Principles and Recommendations for a Vital Statistics System Revision 3. (2014).

34. Kotabagi, R. B., Chaturvedi, R. K. & Banerjee, A. Medical certification of cause of death. Medical Journal Armed Forces India 60, 261–272 (2004).

35. Thacker, S. B. & Berkelman, R. L. Public Health Surveillance in the United States. Epidemiologic Reviews 10, 164–190 (1988).