Social disadvantage, economic inequality, and life expectancy in nine Indian states

Sangita Vyas[*] Payal Hathi[†] Aashish Gupta[‡]

May 4, 2021

[*]University of Texas at Austin and r.i.c.e., sangita.vyas@utexas.edu. 2225 Speedway, Austin, TX 78712.
[†]University of California, Berkeley and r.i.c.e., phathi@berkeley.edu.
[‡]University of Pennsylvania and r.i.c.e., aashishg@sas.upenn.edu.
For helpful comments and suggestions, we are grateful to Monica Alexander, Samuel Arenberg, Casey Breen, Karthik Rao Cavale, Diane Coffey, Irma Elo, Dennis Feehan, Nathan Franz, Michael Geruso, Sonal Gihara-Sharma, Michel Guillot, Rajeev Kumar, Leigh Linden, Nazar Khalid, Kanika Sharma, Dean Spears, Nikkil Sudharsanan, Megan Reed, Amit Thorat, Atheendar Venkataramani, and Tom Vogl. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1610403. This research was also supported by grant P2CHD042849, Population Research Center, and by the grant T32HD007081, Training Program in Population Studies, both awarded to the Population Research Center at The University of Texas at Austin by the Eunice Kennedy Shriver National Institute of Child Health and Human Development, and by a National Institute of Child Health and Human Development Training grant at the University of California, Berkeley, T32HD007275. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the official views of the NSF or the NIH. Additional support was provided by the IUSSP CRVS Fellowship.

Abstract

An extensive literature documents the contributions of discrimination and social exclusion to health disparities. This study investigates life expectancy differentials along lines of caste, religion, and indigenous identity in India, home to some of the largest populations of marginalized social groups in the world. Using a large, high-quality survey that measured mortality, social group, and economic status, we are the first to estimate and decompose life expectancy differences between higher-caste Hindus and three of India's most disadvantaged social groups: Adivasis, Dalits, and Muslims. Relative to higher-caste Hindus, Adivasi life expectancy is more than four years lower, Dalit life expectancy is more than three years lower, and Muslim life expectancy is about one year lower. Economic status explains less than half of these gaps. The differences between the life expectancy of higher-caste Hindus and the life expectancies of Adivasis and Dalits are comparable to the Black-White gap in the US in absolute magnitude. The differences are larger in relative terms because overall life expectancy in India is lower. Our findings extend the literature on fundamental causes of global health disparities. Methodologically, we contribute to the literature on mortality estimation and demographic decomposition using survey data from low- and middle-income contexts.

Keywords: social inequality, life expectancy, India, caste, religion, indigenous identity

Social disadvantage and health are closely linked. In the United States, for example, disparities in health and mortality between Black and White Americans have persisted over decades despite changes in technology, exposures, and diseases (Hayward and Heron, 1999; Williams, Lawrence and Davis, 2019; Phelan and Link, 2015; Wrigley-Field, 2020). However, the health impacts of social exclusion remain severely understudied in low- and middle-income countries (LMICs) (Deaton, 2002). Understanding health disparities in LMICs is important in part because social marginalization is no less present in poorer societies than in rich ones like the United States. Moreover, compared to high-income countries (HICs), patterns of disparities may be distinct in LMICs because population health is poorer, social safety nets are less robust, health care is less accessible, and mortality risk-factors differ (Tabutin and Masquelier, 2017).

This article describes and decomposes life expectancy disparities between socially marginalized and privileged groups in one of the most populated and stratified countries in the world, India. The social groups we study – Dalits, Adivasis, and Muslims – experience social exclusion based on caste, indigenous identity, or religion, respectively (Ambedkar, 1937; Xaxa, 1999; Sachar Commitee, Government of India, 2006). Although each social group faces distinct forms of marginalization, together they comprise a population of over 450 million, greater than that of the US. Each group individually is also among the largest marginalized social groups in the world. We compare mortality for these three groups to that of higher-caste Hindus, who are relatively privileged in Indian society.

As in other contexts, social and economic disadvantages occur simultaneously in India. Dalits, Adivasis, and Muslims are poorer than privileged groups (Desai and Kulkarni, 2008; Azam and Bhatt, 2015). To quantify the extent to which differences in socioeconomic status (SES) can explain mortality differences between groups, such as between Black and White Americans (Williams and Collins, 1995; Geruso, 2012), the scientific literature has relied on standardization and decomposition techniques. These analyses are made possible in HICs by complete vital registration and multiple large surveys that reliably measure mortality, race, and SES (Elo, 2009; Warren et al., 2017; Brown, Lariscy and Kalousová, 2019). In many LMICs, however, decomposing life expectancy differences between groups is constrained by the sparseness of data on all-cause mortality linked with social conditions (Clark, 2019; Saikia and Kulkarni, 2016). For this reason, direct estimation of life tables disaggregated by social group and SES is not possible.

We overcome this limitation by using a unique and large-scale survey in nine Indian states in 2010-2011 that collected retrospective mortality information, social group, and SES from four million households. The sample is sufficient for directly estimating age-specific mortality rates. We find that the overall age-specific mortality rates estimated from this data correspond closely to official life tables. In particular, we compare our sex-specific, aggregated life tables to the life tables generated by the Government of India's Sample Registration System (SRS), a nationally representative system of mortality monitoring that does not disaggregate data by social group. After establishing the credibility of overall mortality estimates, we construct period sex-, group- and SES-specific life tables using standard demographic approaches. To examine the extent to which differences in SES between groups account for differences in life expectancy, we use a non-parametric standardization technique (DiNardo, Fortin and Lemieux, 1996; Geruso, 2012; Coffey, 2015). We use a cluster-bootstrap strategy to calculate standard errors for the life table and decomposition quantities we estimate (Cameron and Miller, 2015).

We are the first to document lower life expectancy at birth among Adivasis and Dalits compared to higher-caste Hindus within each state. Relative to higher-caste Hindus, Adivasi life expectancy at birth is about four years lower for females and five years lower for males. Life expectancy gaps between higher-caste Hindus and Dalits is more than three years. We also provide the first estimates of Muslim life expectancy in India. In this region, Muslim life expectancy at birth is about one year lower than it is for higher-caste Hindus. We find that lower life expectancy for all three groups relative to higher-caste Hindus is not fully explained by differences in SES. Most of the gaps remain after accounting for differences in rural residence, wealth, and environmental factors.

Our findings have implications for health disparities worldwide, as well as for the global burden of mortality. The estimates of life expectancy at birth for Adivasis and Dalits are comparatively low globally. They are, for example, lower than contemporaneous population-level life expectancy at birth in many poorer contexts in sub-Saharan Africa. In terms of years, the disadvantages that we estimate for Adivasis and Dalits relative to higher-caste Hindus are comparable to the Black-White gap in the US. They are also comparable to the gap between the US and other HICs during the same time period. In percentage terms, the disparities we observe are more substantial because life expectancy in India is less than four-fifths the level of life expectancy in the US. Compared to existing estimates on the extent to which SES accounts for the Black-White life expectancy gap (Geruso, 2012), the SES factors we use here account for less of the gaps between marginalized and privileged groups. Because the population we study is so large, the number of life-years lost because of lower life expectancy for Adivasis and Dalits compared to higher-caste Hindus in this sample is almost four times the number of life-years lost because of the life expectancy disadvantage for Black Americans relative to White Americans.

Our paper makes several contributions to the literature on social disadvantage and health in LMICs. First, we advance the scientific study of estimating mortality in LMICs (Masquelier, 2013; Helleringer et al., 2014; Feehan and Borges, 2019), which has been stymied by the lack of reliable and direct population-representative estimates (Hill, 1991; Moultrie et al., 2013). We document an example in which a retrospective question on deaths in the household in the recent period produces estimates that follow typical patterns of mortality across the life course. These estimates also match expected levels of age-specific death rates in this context. Second, by examining the extent to which economic status can account for life expectancy disparities, our paper contributes to the existing literature using standardization and decomposition methods (Burgard, 2002; Dommaraju, Agadjanian and Yabiku, 2008; Bora, Lutz and Raushan, 2018; Coffey et al., 2019; Victora et al., 2020). Finally, we further the methodological literature on statistical inference of life table estimates constructed using survey data (Feehan, Mahy and Salganik, 2017; Chiu, Hayward and Saito, 2016; Cai et al., 2010; Wakefield et al., 2019).

From a policy perspective, the Indian constitution provides protections for certain marginalized groups. However, discussions on health and well-being within and outside India often ignore inequalities based on indigenous identity, caste, and religion. This study highlights the importance of measuring and addressing social disparities within India and other LMIC contexts.

2 Data

This study uses the Government of India's Annual Health Survey (AHS) 2010-2011, a household survey that visited over 4 million households across approximately 20,000 primary sampling units (PSUs), and collected data on over 20 million individuals in nine relatively poor states in India: Assam, Bihar, Chhattisgarh, Jharkhand, Madhya Pradesh, Odisha, Rajasthan, Uttar Pradesh, and Uttarakhand. These states represent 48.5% of India's population (Office of the Registrar General, 2011). We use data from the household roster, which recorded information on usual members living in surveyed households on January 1, 2010. We also use data from the mortality roster, which recorded the characteristics of surveyed households' usual members that died between January 1, 2007 and December 31, 2009. AHS instructions to enumerators are described in Appendix Table A1. Appendix Table A2 describes the sample.

The AHS recorded the social group and religion of each household. We focus on estimating life expectancies for India's largest social groups: Dalits, Adivasis, Muslims, and Other Backward Caste (OBC)/high-caste Hindus. The data do not allow us to separately identify OBCs and high-caste Hindus. OBCs are relatively priveleged compared to Dalits and Adivasis but marginalized compared to high-caste Hindus. Therefore, we expect the combined life expectancy of OBCs and high-castes to be lower than it would be for high-caste Hindus on their own (Gupta and Coffey, 2020). Given that the OBC population is greater than that of high-caste Hindus (Desai et al., 2010), the combined life expectancy figure is likely to be closer to that of OBCs than that of high-castes. Since OBCs are intermediate caste groups (Galanter, 1978), we refer to the combined OBC and high-caste group as either higher-caste Hindus or OBC/high-caste Hindus in this article. Further details on measurement of social group are in Appendix A2.1.

The AHS also recorded data on household SES, including rural residence, wealth, and environmental exposures. To summarize a household's wealth, we construct a wealth index using a principal component analysis of asset ownership and house infrastructure. The index is described in greater detail in Appendix A2.2. The relative disadvantage of Adivasis, Dalits, and Muslims, compared to OBC/high-caste Hindus, is evident in Appendix Figure A1 and Table A3, which display summary statistics by social group. Data are described further in Appendix A2.

3 Methods

3.1 Estimating mortality rates and life expectancy

Using data from the AHS household and mortality rosters, we construct a dataset that records the number of person-years each individual contributes to each single-year age during the period January 2007 through December 2009, and whether the individual died at that age. We estimate social group-, age-, and sex-specific mortality rates (shown in Appendix Figure A2). Using standard procedures, we construct eight life tables to calculate life expectancy at birth separately for males and females of each social group. The number of person-years lived by those who died in each age interval, or ${}_na_x$, is calculated based on values from the Government of India's SRS 2007-2011 official life tables for states. In additional analyses, we estimate social group-, state-, and sex-specific life expectancies, as well as social group-, wealth-decile-, and sex-specific ones using the same methods. All estimates use the sample weights provided in the survey to make the data representative of the nine AHS states. Methods for estimating life expectancy are described further in Appendix A1.1.

In order to rule out concerns regarding data quality from retrospective survey questions about mortality within the household (Hill, 1991; Timaeus, 1991), we compare age-specific mortality rates estimated from the AHS to those from the SRS and the the National Family Health Survey (NFHS), India's Demographic and Health Survey (Appendix A2.3 and Appendix Table A1 describe the SRS and NFHS in greater detail). Figure 1 displays this analysis. Age-specific rates estimated from the AHS match closely with rates from the SRS and NFHS. They are also smoother than those from SRS and NFHS. Effects of age-misreporting, which are known to be high in India (Gerland, 2014), are apparent at age 75 in the AHS. This is also present to some extent in the SRS and NFHS surveys at older ages. Appendix Figure A3 shows that unadjusted mortality rates are similar to Gompertz rates (in ages 40-85+), which adjust for age-misreporting.

This age-misreporting, which is more common among more disadvantaged groups, may lead to downward biases in estimates of mortality (Preston, Elo and Stewart, 1999). The disparities in life expectancy between marginalized groups and higher-caste Hindus that we document are likely conservative for at least two more reasons. First, because we cannot distinguish between high-caste Hindus and OBCs, we expect estimates of life expectancy for the combined higher-caste Hindu group to represent a lower bound for high-caste Hindu life expectancy. Second, we are not able to capture mortality that occurs in households in which all members have died, which is more likely in Adivasi and Dalit households. We expect this particular bias to be relatively small. If mortality in single-person households is similar to mortality in two-person households, the closest counterfactual in our data, the number of deaths the AHS would have missed is less than 1% of total observed deaths.

3.2 Demographic Re-weighting

To understand the extent to which differences in SES can account for social-group differentials in life expectancy, we use a demographic re-weighting strategy (DiNardo, Fortin and Lemieux, 1996; Geruso, 2012). Non-parametric re-weighting techniques allow for studying non-linear functions like life expectancy and, by matching on the full distribution of observed characteristics, are more flexible than regression techniques.

In practice, the demographic re-weighting technique estimates counterfactual life expectancies for marginalized social groups, re-weighting these groups so that they match the distribution of SES among OBC/high-caste Hindus. Because the marginalized social groups we study have lower SES than OBC/high-caste Hindus, the re-weighting strategy produces counterfactual life expectancies by up-weighting wealthier individuals and down-weighting less wealthy individuals in each marginalized group. This is implemented by estimating a re-weighting function as follows:

$$\psi^{MG}(c_i) = \frac{f(c_i|HC)}{f(c_i|MG)}, \tag{1}$$

where MG represents the marginalized group considered, and HC represents OBC/high-caste Hindus. c_i represents a vector of observable characteristics for individual i that are correlated with life expectancy, including sex, age group, social group, and SES. f represents the probability density function. Each individual in the sample is multiplied by her corresponding re-weighting function to produce counterfactual age-specific mortality rates and counterfactual life expectancies for each marginalized group. Re-weighting methods are described further in Appendix A1.2.

The SES characteristics included in the re-weighting exercise are determined based on regression analysis of the characteristics associated with mortality (see Appendix Tables A4 and A5). They include rural residence, wealth (the intersection of wealth index quintile and land ownership), and environmental exposures (household solid fuel use for individuals age five and older, and household solid fuel use intersected with four categories of the fraction of people defecating in the open in the PSU for children younger than age five). These factors have also been identified in the prior literature as important determinants of mortality (Link and Phelan, 1995; Smith, 2000; Spears, 2018). The regression equations are described further in Appendix A1.3.

3.3 Inference: Cluster-bootstrap Method

Standard errors are estimated using the cluster-boostrap method described in Cameron and Miller (2015). We use this procedure because the AHS randomly sampled PSUs (villages or census enumeration blocks) rather than individuals, and both outcomes and explanatory variables are likely correlated within PSUs. Chiu, Hayward and Saito (2016) and Cai et al. (2010) similarly construct standard errors around life table quantities estimated from cluster sample surveys.

Within districts, the AHS randomly sampled villages and urban areas stratified by population size. The number of PSUs randomly sampled from each stratum were determined based on the district's population distribution across strata. All households in sampled PSUs were interviewed.

For the bootstrap, we resample with replacement $J_{strat,dist}$ PSUs within each district-stratum, with $J_{strat,dist}$ equal to the total number of PSUs in that district's stratum in the original AHS sample. Because our resampling procedure maintains the distribution of PSUs across strata within districts, we use the original AHS sample weights, which vary at the district-stratum level, to analyse each resample. Using the dataset generated by each resample, we estimate age-specific mortality rates and life tables. For the decomposition, we estimate a new re-weighting function and counterfactual life tables using each resample. We repeat this process 500 times, and the standard deviation of the 500 resulting estimates for each statistic are used for calculating 95% confidence intervals.

4 Results

4.1 Marginalized social groups have lower life expectancies

Figure 2 shows female and male life expectancies at birth for the four social groups we study. It also shows 95% confidence intervals calculated using a cluster-bootstrap approach. Compared to high-caste Hindus/OBCs, we observe lower life expectancies at birth among marginalized social groups. Adivasis have the lowest life expectancy among the four groups. Differentials between Adivasis and higher-caste Hindus are 3.7 years for women and almost five years for men. The gap between Dalits and higher-caste Hindus is of similar magnitude: more than 3 years for both women and men. Muslim life expectancy is about one year less than that of higher-caste Hindus. Both overall levels of mortality among marginalized social groups and the absolute differentials between groups are comparatively large. Life expectancy for Dalits and Adivasis is similar to those of the poorest countries in the world. The overall gaps are similar in absolute terms to the contemporaneous Black-White gap in the United States (Arias and Xu, 2018) and the Arab-Jewish gap in Israel (Saabneh, 2016)

The life expectancy advantage for women is highest among Adivasis and lowest among higher-caste Hindus. This pattern is noteworthy because Adivasis have more egalitarian gender norms than other groups (Maharatna, 2000).

That Muslims have overall lower life expectancy compared to higher-caste Hindus, but that the gaps are smaller relative to other marginalized groups, is also consistent with the prior literature (Sachar Commitee, Government of India, 2006; Guillot and Allendorf, 2010; Bhalotra, Valente and Van Soest, 2010). This literature has identified lower mortality among Muslim children because of lower exposure to open defecation (Geruso and Spears, 2018). Muslims also have lower rates of cervical cancers (Gomes et al., 2017), alcohol consumption (International Institute for Population Sciences - IIPS/India and ICF., 2017), and suicide (Thimmaiah et al., 2016; Lester, 2006).

Research on caste and social identity has emphasized that although there are features of social stratification that are common across India, marginalization manifests differently from region to region (Srinivas et al., 1962). Mortality risks also vary across states. Figure 3 shows life expectancy by social group, sex, and state. We do not estimate mortality rates for social groups that constitute less than five percent of a state's population. We find that across states, Dalits and Adivasis have lower life expectancy compared to higher-caste Hindus. Except for one state, Muslims have similar or lower life expectancy than higher-caste Hindus. Among the nine states, Adivasi life expectancy is highest in Assam, a society in which they face less discrimination compared to other AHS sample states (Gogoi and Saikia, 2020). The life expectancy of Dalits is lowest in Uttar Pradesh, and that of Adivasis is lowest in Madhya Pradesh.

4.2 Life expectancy disparities remain after accounting for SES

Given that marginalized social groups are also poorer, to what extent are social group differences in life expectancy driven by economic disadvantage? Figure 4 shows life expectancy by social group and household wealth decile. As with state patterns, we find that Adivasis and Dalits have lower life expectancies than higher-caste Hindus across wealth categories. Among the bottom deciles, Muslim life expectancy is similar to higher-caste life expectancy. However, among wealthier deciles, Muslims have lower life expectancy. These patterns for Muslims are consistent with explanations related to alcohol consumption and associated mortality. Because of notions related to caste and sanskritization (Srinivas, 1956), alcohol consumption among richer higher castes has also been shown to be low (Subramanian et al., 2005).

Figure 5 explores the extent to which dimensions additional to household wealth can explain differences in life expectancy between social groups. It uses the re-weighting approach outlined in Section 3.2. We find that differences in rural residence, wealth, and environmental exposures do not fully account for the life expectancy gaps between marginalized social groups and higher-caste Hindus.

The figure shows gaps in life expectancy at birth between marginalized groups and higher-caste Hindus, separately for females and males. The vertical lines in the figure reflect 95% confidence intervals. The left-most estimates show the raw gaps. The differences are about one year for Muslim men and women; more than three years for Dalit women, Adivasi women, and Dalit men; and about five years for Adivasi men. The second set of estimates shows the gaps that remain after re-weighting the marginalized groups to reflect the distribution across rural and urban residence among higher-caste Hindus. Accounting for rural residence reduces the gap for Adivasis and Dalits, but not substantially. For Muslims, it increases the gap slightly, because Muslims are more likely to live in urban areas than higher-caste Hindus. The third set of estimates adds wealth quintile and land ownership to the re-weighting characteristics. Except for Muslim women, this reduces gaps. The right-most set of estimates adds environmental exposures. For children under age five, these include household solid fuel use and the fraction of individuals defecating in the open in the PSU. For individuals age five and older, we only include household solid fuel use. Environmental exposures do not explain gaps substantially, and for Muslims, they actually increase gaps, given lower exposure to open defecation among Muslims relative to higher-caste Hindus (Geruso and Spears, 2018).

In the United States, observable SES characteristics explain about three-quarters of the life expectancy gap between Black and White Americans (Geruso, 2012). In India, they explain less than half. We find that the unexplained gaps are about half a year for Muslim men, one year for Muslim women, more than two years for Adivasi women, Dalit women, and Dalit men, and more than three years for Adivasi men.

5 Discussion

This study is the first to examine relationships between social disadvantage, economic status, and life expectancy in India. Using a novel survey, we document large and important disadvantages in life expectancy at birth for Adivasis, Dalits, and Muslims compared to higher-caste Hindus. These disparities cannot be explained by differences in rural residence, wealth, or environmental exposures.

Life expectancies for Adivasis and Dalits are some of the lowest in the world. In terms of population, these groups are also larger than most countries in the world. Our study therefore underscores the global significance of challenging social inequality in health in India. Indeed, extreme social stratification and exploitation in India may be contributing to global population health deficits and slower improvements in health worldwide.

From a global comparative perspective, life expectancy gaps in India are similar in magnitude to ethnic and identity-based disparities such as by race in the US, Brazil, and South Africa, indigenous identity in New Zealand and Australia, and religion in Israel (Arias and Xu, 2018; Chiavegatto Filho, Beltrán-Sánchez and Kawachi, 2014; Gupta and Sudharsanan, 2020; Phillips et al., 2017; Saabneh, 2016). Even within India, the disparities in life expectancy between Adivasis and Dalits on the one hand, and higher-caste Hindus on the other, are large. Life expectancy for higher-caste Hindus in the nine AHS states is higher than the contemporaneous all-India life expectancy observed around 2010 (Registrar General and Census Commissioner of India, 2010). Life expectancies for Adivasis and Dalits, however, are lower than those observed for all of India in 1996-2000, more than ten years before the AHS survey (Registrar General and Census Commissioner of India, 2013). The gaps documented in this paper are also larger than the within-sample gap in life expectancy at birth between individuals at the 25th and 75th percentiles of the wealth distribution.

Our findings highlight several first-order concerns for future research. Given ongoing epidemiological transitions in India and other LMICs, continuous monitoring of mortality within countries and globally is important. This study documents the value of several approaches for studying mortality and its determinants in LMICs. These include large-scale data collection exercises that contain retrospective questions on household deaths, empirical estimation of age-specific mortality rates, non-parametric re-weighting techniques, and cluster-bootstrap variance estimation approaches suitable for multi-stage sample survey data. Research on behaviors, age-specific contributions, causes of death, and risk-factors, such as occupational exposures, may help further understand disparities in mortality and life expectancy. Following the qualitative and auto-biographical literature (Kamble, 1990; Valmiki and Mukherjee, 2008; Shah et al., 2006; Acharya, 2010; George, 2019), studies that are able to document causal pathways from violence, exploitation, and discrimination to mortality are also valuable.

From a policy perspective, these findings suggest that population health interventions that explicitly challenge social disadvantage are essential because addressing economic inequality may not be sufficient. Unfortunately, health policy in India and globally largely ignores caste, religion, exploitation, and discrimination. This study justifies further action on social disparities in health within India, and advances the global conversation addressing inequalities based on race, ethnicity, indigenous identity, caste, and religion.

References

Acharya, Sanghmitra. 2010. "Caste and patterns of discrimination in rural public health care services." *Blocked by Caste: Economic Discrimination in Modern India (New Delhi: Oxford University Press, 2010)*, 208–229.

Ambedkar, Bhimrao Ramji. 1937. *Annihilation of caste: With a reply to Mahatama Gandhi.* University of Columbia.

Arias, Elizabeth, and Jiaquan Xu. 2018. "United States life tables, 2015."

Azam, Mehtabul, and Vipul Bhatt. 2015. "Like father, like son? Intergenerational educational mobility in India." *Demography*, 52(6): 1929–1959.

Bhalotra, Sonia, Christine Valente, and Arthur Van Soest. 2010. "The puzzle of Muslim advantage in child survival in India." *Journal of Health Economics*, 29(2): 191–204.

Bora, Jayanta Kumar, Wolfgang Lutz, and Rajesh Raushan. 2018. "Contribution of education to infant and under-five mortality disparities among caste groups in India." Vienna Institute of Demography Working Papers.

Brown, Dustin C, Joseph T Lariscy, and Lucie Kalousová. 2019. "Comparability of mortality estimates from social surveys and vital statistics data in the United States." *Population research and policy review*, 38(3): 371–401.

Burgard, Sarah. 2002. "Does race matter? Children's height in Brazil and South Africa." *Demography*, 39(4): 763–790.

Cai, Liming, Mark D Hayward, Yasuhiko Saito, James Lubitz, Aaron Hagedorn, and Eileen Crimmins. 2010. "Estimation of multi-state life table functions and their variability from complex survey data using the SPACE Program." *Demographic research*, 22(6): 129.

Cameron, A Colin, and Douglas L Miller. 2015. "A practitioner's guide to cluster-robust inference." *Journal of Human Resources*, 50(2): 317–372.

Chiavegatto Filho, Alexandre Dias Porto, Hiram Beltrán-Sánchez, and Ichiro Kawachi. 2014. "Racial disparities in life expectancy in Brazil: challenges from a multiracial society." *American journal of public health*, 104(11): 2156–2162.

Chiu, Chi-Tsun, Mark Hayward, and Yasuhiko Saito. 2016. "a comparison of educational differences on physical health, mortality, and healthy life expectancy in Japan and the United States." *Journal of aging and health*, 28(7): 1256–1278.

Clark, Samuel J. 2019. "A General Age-Specific Mortality Model With an Example Indexed by Child Mortality or Both Child and Adult Mortality." *Demography*, 56(3): 1131–1159.

Coffey, Diane. 2015. "Prepregnancy body mass and weight gain during pregnancy in India and sub-Saharan Africa." *Proceedings of the National Academy of Sciences*, 112(11): 3302–3307.

Coffey, Diane, Ashwini Deshpande, Jeffrey Hammer, and Dean Spears. 2019. "Local Social Inequality, Economic Inequality, and Disparities in Child Height in India." *Demography*, 56(4): 1427–1452.

Deaton, Angus. 2002. "Commentary: The convoluted story of international studies of inequality and health." *International Journal of Epidemiology*, 31(3): 546–549.

Desai, Sonalde, and Veena Kulkarni. 2008. "Changing educational inequalities in India in the context of affirmative action." *Demography*, 45(2): 245–270.

Desai, Sonalde B, Amaresh Dubey, Brij Lal Joshi, Mitali Sen, Abusaleh Shariff, and Reeve Vanneman. 2010. "Human development in India." *New York: Oxford University.*

DiNardo, John, Nicole M Fortin, and Thomas Lemieux. 1996. "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach." *Econometrica*, 64(5): 1001–1044.

Dommaraju, Premchand, Victor Agadjanian, and Scott Yabiku. 2008. "The pervasive and persistent influence of caste on child mortality in India." *Population Research and Policy Review*, 27(4): 477–495.

Elo, Irma T. 2009. "Social class differentials in health and mortality: Patterns and explanations in comparative perspective." *Annual review of sociology*, 35: 553–572.

Feehan, Dennis M, and Gabriel M Borges. 2019. "Estimating adult death rates from sibling histories: A network approach." *arXiv preprint arXiv:1906.12000.*

Feehan, Dennis M, Mary Mahy, and Matthew J Salganik. 2017. "The network survival method for estimating adult mortality: Evidence from a survey experiment in Rwanda." *Demography*, 54(4): 1503–1528.

Filmer, Deon, and Lant H Pritchett. 2001. "Estimating wealth effects without expenditure data—or tears: an application to educational enrollments in states of India." *Demography*, 38(1): 115–132.

Galanter, Marc. 1978. "Who are the other backward classes?: An introduction to a constitutional puzzle." *Economic and Political Weekly*, 1812–1828.

George, Sobin. 2019. "Reconciliations of Caste and Medical Power in Rural Public Health Services." *Economic and Political Weekly*, 54(40).

Gerland, Patrick. 2014. "UN Population Division's methodology in preparing base population for projections: case study for India." *Asian Population Studies*, 10(3): 274–303.

Geruso, Michael. 2012. "Black-white disparities in life expectancy: how much can the standard SES variables explain?" *Demography*, 49(2): 553–574.

Geruso, Michael, and Dean Spears. 2018. "Neighborhood sanitation and infant mortality." *American Economic Journal: Applied Economics*, 10(2): 125–62.

Gogoi, Suraj, and Parag Jyoti Saikia. 2020. "Sharing and exchanging: understanding common grounds in northeast India." *Asian Ethnicity*, 21(1): 22–38.

Gomes, Mireille, Rehana Begum, Prabha Sati, Rajesh Dikshit, Prakash C Gupta, Rajesh Kumar, Jay Sheth, Asad Habib, and Prabhat Jha. 2017. "Nationwide mortality studies to quantify causes of death: relevant lessons from India's Million Death Study." *Health Affairs*, 36(11): 1887–1895.

Guillot, Michel, and Keera Allendorf. 2010. "Hindu-Muslim differentials in child mortality in India." *Genus*, 66(2).

Gupta, Aashish, and Diane Coffey. 2020. "Caste, religion, and mental health in India." *Population Research and Policy Review*, 1–23.

Gupta, Aashish, and Nikkil Sudharsanan. 2020. "Large and persistent life expectancy disparities among India's social groups."

Hayward, Mark D, and Melonie Heron. 1999. "Racial inequality in active life among adult Americans." *Demography*, 36(1): 77–91.

Helleringer, Stéphane, Gilles Pison, Almamy M Kanté, Géraldine Duthé, and Armelle Andro. 2014. "Reporting errors in siblings' survival histories and their impact on adult mortality estimates: results from a record linkage study in Senegal." *Demography*, 51(2): 387–411.

Hill, Kenneth. 1991. "Approaches to the measurement of childhood mortality: a comparative review." *Population Index*, 57(3): 368–382.

International Institute for Population Sciences - IIPS/India and ICF. 2017. *India National Family Health Survey NFHS-4 2015-16.* Mumbai, India: IIPS and ICF.

Kamble, Baby. 1990. "Jina Amucha." *Pune: Mansanman Prakashan. 2007 The Prisons We Broke.*

Lester, David. 2006. "Suicide and islam." *Archives of suicide research*, 10(1): 77–97.

Link, Bruce G, and Jo Phelan. 1995. "Social conditions as fundamental causes of disease." *Journal of health and social behavior*, 80–94.

Maharatna, Arup. 2000. "Fertility, mortality and gender bias among tribal population: an Indian perspective." *Social Science & Medicine*, 50(10): 1333–1351.

Masquelier, Bruno. 2013. "Adult mortality from sibling survival data: a reappraisal of selection biases." *Demography*, 50(1): 207–228.

Moultrie, Tom A, RE Dorrington, Allan G Hill, Kenneth Hill, IM Timæus, and Basia Zaba. 2013. *Tools for demographic estimation.* International Union for the Scientific Study of Population.

Office of the Registrar General. 2011. "Census of India 2011: Household Listing Operations." *Office of the Registrar and General Census Commissioner, India. Indian Census Bureau.*

Phelan, Jo C, and Bruce G Link. 2015. "Is racism a fundamental cause of inequalities in health?" *Annual Review of Sociology*, 41: 311–330.

Phillips, Bronwen, John Daniels, Alistair Woodward, Tony Blakely, Richard Taylor, and Stephen Morrell. 2017. "Mortality trends in Australian Aboriginal peoples and New Zealand Māori." *Population health metrics*, 15(1): 25.

Preston, Samuel H, Irma T Elo, and Quincy Stewart. 1999. "Effects of age misreporting on mortality estimates at older ages." *Population studies*, 53(2): 165–177.

Preston, S, Patrick Heuveline, and Michael Guillot. 2000. "Demography: measuring and modeling population processes. 2001." *Malden, MA: Blackwell Publishers.*

Registrar General and Census Commissioner of India. 2010. *Sample Registration System Technical Report 2010.* Government of India.

Registrar General and Census Commissioner of India. 2013. *Compendium of India's Fertility and Mortality Indicators ,1971 - 2013.* Government of India.

Registrar General and Census Commissioner of India. 2020. *Sample Registration System Statistical Report 2018.* Government of India.

Saabneh, Ameed M. 2016. "Arab–Jewish gap in life expectancy in Israel." *The European Journal of Public Health*, 26(3): 433–438.

Sachar Commitee, Government of India. 2006. "Social, Economic and Educational Status of the Muslim Community of India." *Prime Minister's High Level Committee, Cabinet Secretariat, Government of India.*

Saikia, Nandita, and PM Kulkarni. 2016. "Data for Research into Health Inequality in India." *The Economic and Political Weekly*, 51(26): 111–116.

Shah, G., H. Mander, S. Thorat, A. Baviskar, and S. Deshpande. 2006. *Untouchability in Rural India.* SAGE Publications.

Smith, Kirk R. 2000. "National burden of disease in India from indoor air pollution." *Proceedings of the National Academy of Sciences*, 97(24): 13286–13293.

Spears, Dean. 2018. "Exposure to open defecation can account for the Indian enigma of child height." *Journal of Development Economics.*

Srinivas, Mysore Narasimhachar. 1956. “A note on Sanskritization and Westernization.” *The Far Eastern Quarterly*, 15(4): 481–496.

Srinivas, Mysore Narasimhachar, et al. 1962. “Caste in modern India and other essays.” *Caste in modern India and other essays.*

Subramanian, SV, Shailen Nandy, Michelle Irving, David Gordon, and George Davey Smith. 2005. “Role of socioeconomic markers and state prohibition policy in predicting alcohol consumption among men and women in India: a multilevel statistical analysis.” *Bulletin of the World Health Organization*, 83: 829–836.

Tabutin, Dominique, and Bruno Masquelier. 2017. “Mortality inequalities and trends in low-and middle-income countries, 1990-2015.” *Population*, 72(2): 221–296.

Tai, Tzu Han, and Andrew Noymer. 2018. “Models for estimating empirical Gompertz mortality: With an application to evolution of the Gompertzian slope.” *Population Ecology*, 60(1-2): 171–184.

Thimmaiah, Rohini, Vijayalakshmi Poreddi, Rajalakshmi Ramu, Sugavana Selvi, and Suresh Bada Math. 2016. “Influence of religion on attitude towards suicide: An Indian perspective.” *Journal of religion and health*, 55(6): 2039–2052.

Timaeus, Ian M. 1991. “Measurement of adult mortality in less developed countries: a comparative review.” *Population index*, 552–568.

UN-IGME. 2018. “Levels and trends in child mortality: 2017 Report.” UNICEF-WHO-WB-UNPD. http://www. childmortality. org. Accessed 22 Apr.

Valmiki, Omprakash, and Arun Prabha Mukherjee. 2008. *Joothan: An Untouchable’s Life.* Columbia University Press.

Victora, Cesar G, Aluisio JD Barros, Cauane Blumenberg, Janaina Calu Costa, Luis Paulo Vidaletti, Fernando C Wehrmeister, Bruno Masquelier, Lucia Hug, and Danzhen You. 2020. “Association between ethnicity and under-5 mortality: analysis of data from demographic surveys from 36 low-income and middle-income countries.” *The Lancet Global Health*, 8(3): e352–e361.

Wakefield, Jon, Geir-Arne Fuglstad, Andrea Riebler, Jessica Godwin, Katie Wilson, and Samuel J Clark. 2019. “Estimating under-five mortality in space and time in a developing world context.” *Statistical methods in medical research*, 28(9): 2614–2634.

Warren, John Robert, Carolina Milesi, Karen Grigorian, Melissa Humphries, Chandra Muller, and Eric Grodsky. 2017. “Do inferences about mortality rates and disparities vary by source of mortality information?” *Annals of epidemiology*, 27(2): 121–127.

Williams, David R, and Chiquita Collins. 1995. “US socioeconomic and racial differences in health: patterns and explanations.” *Annual review of sociology*, 21(1): 349–386.

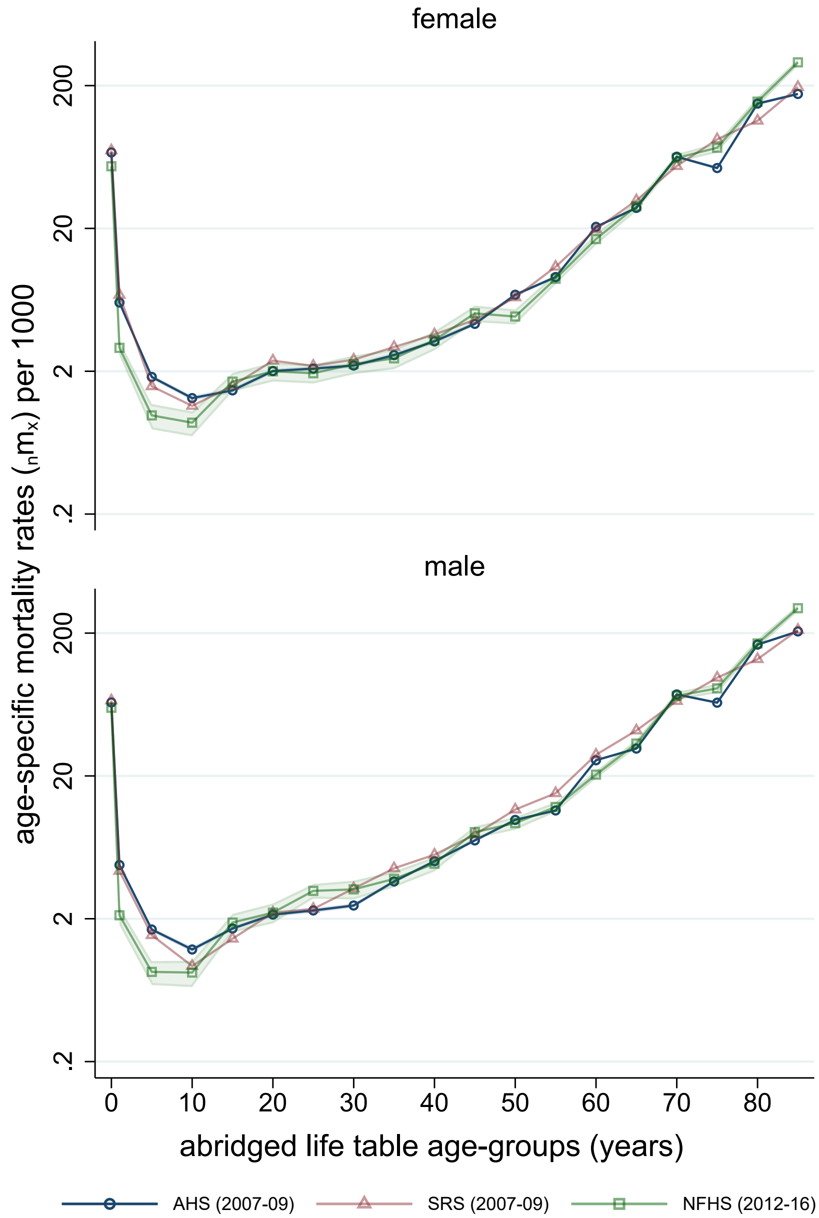
Williams, David R, Jourdyn A Lawrence, and Brigette A Davis. 2019. "Racism and health: evidence and needed research." *Annual review of public health*, 40: 105–125.

World Health Organization. 2016. "WHO Methods and Data Sources for Life Tables 1990–2016." *Global Health Estimates Technical Paper WHO/HIS/IER/GHE/2016.2, WHO, Geneva, May.*

Wrigley-Field, Elizabeth. 2020. "US racial inequality may be as deadly as COVID-19." *Proceedings of the National Academy of Sciences*, 117(36): 21854–21856.

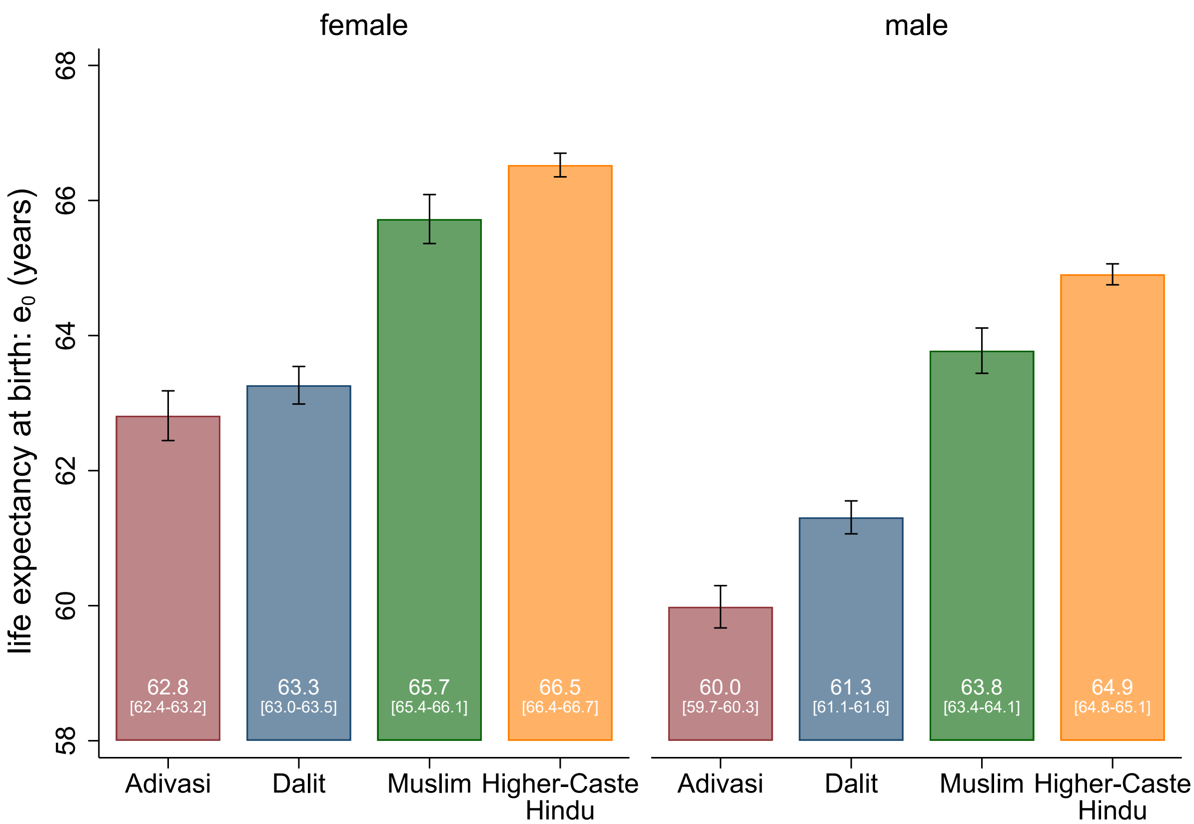
Xaxa, Virginius. 1999. "Tribes as indigenous people of India." *Economic and political weekly*, 3589–3595.

Figure 1: Similar age-specific mortality rates from the Annual Health Survey, Sample Registration System, and National Family Health Survey



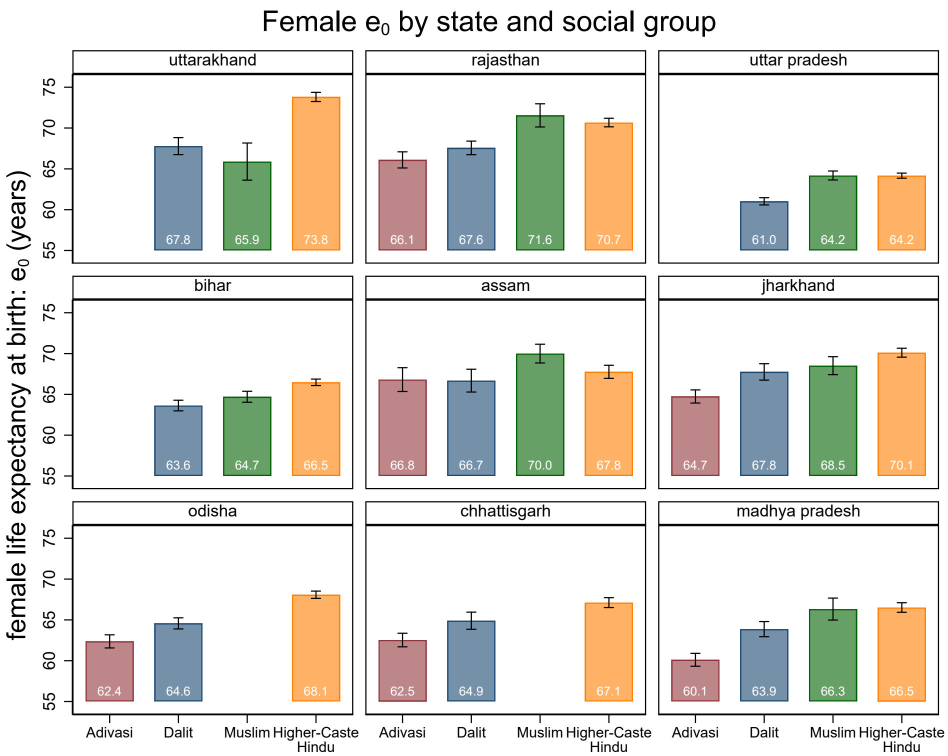
Age-specific mortality rates are estimated for females and males separately, from three different surveys. All AHS states are included, except for Uttarakhand, for which SRS data are missing. For the AHS, mortality rates are estimated based on the procedure described in Section 3.1. NFHS mortality rates are estimated according to the procedure described in Gupta and Sudharsanan (2020). These rates are lower because they are for a later period. For the SRS, we average published age-specific mortality rates for 2007, 2008, and 2009, and weight states by their 2011 Census populations to produce an average across the eight states. Estimates from the AHS and NFHS use sample weights. 95% confidence intervals calculated using a cluster-bootstrap procedure are shown as shaded areas around the AHS (not visible because they are small) and NFHS lines. The SRS does not provide clustered standard errors, so we do not show 95% confidence intervals for SRS rates. Source: Annual Health Survey 2010-2011, Sample Registration System 2007-2009, and National Family Health Survey 2015-2016.

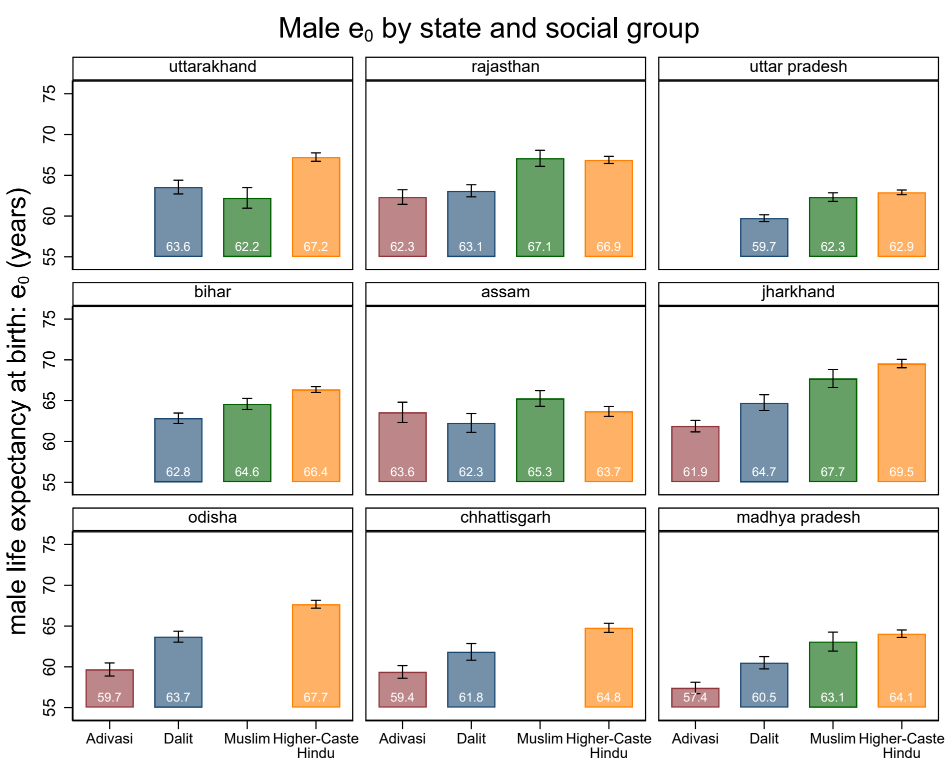
Figure 2: Lower life expectancy at birth among marginalized groups compared to OBC/high-caste Hindus



Life expectancy estimates for each sex and social group are calculated using standard life table procedures. Estimates use sample weights. The vertical lines around each estimate represent 95% confidence intervals calculated using a cluster-bootstrap procedure. Source: Annual Health Survey 2010-2011.

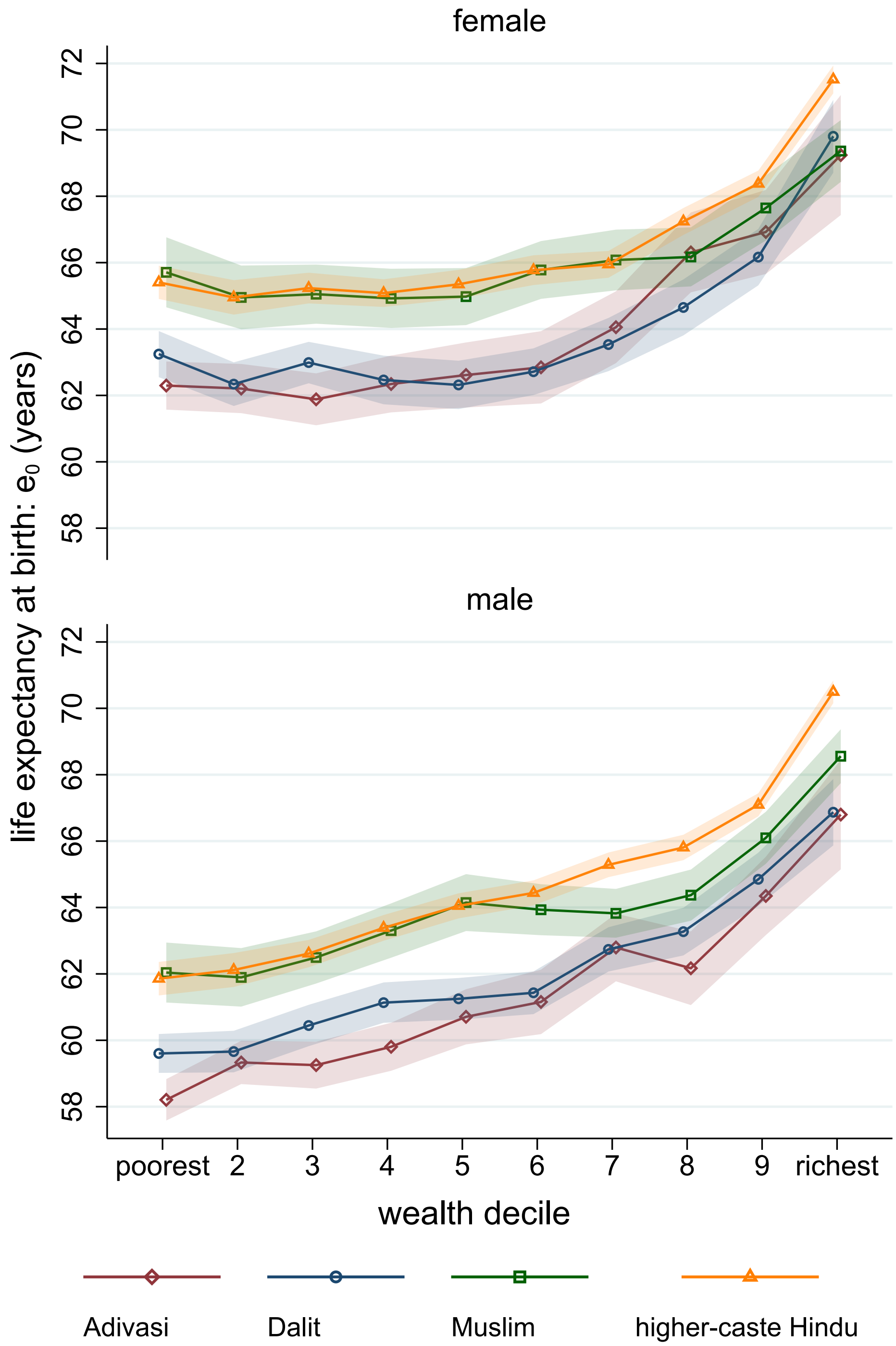
Figure 3: Life expectancy by state: Lower life expectancy at birth among Dalits and Adivasis within states





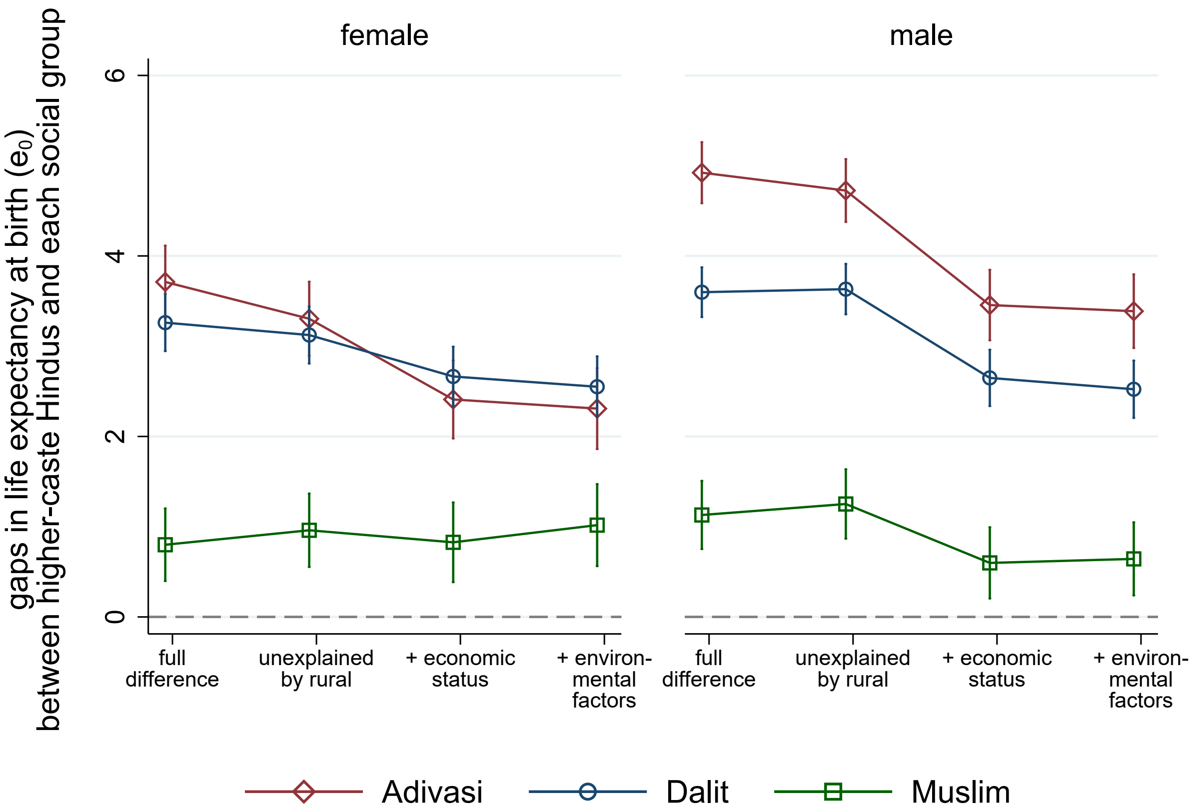
Life expectancy estimates for each state, sex, and social group are calculated using standard life table procedures. Groups which are less than 5% of a state's population are not shown. Estimates use sample weights. The vertical lines around each estimate represent 95% confidence intervals calculated using a cluster-bootstrap procedure. Source: Annual Health Survey 2010-2011.

Figure 4: Life expectancy by household wealth: Lower life expectancy at birth among Dalits and Adivasis at all levels of wealth



Wealth decile are deciles of a wealth index constructed using a principal component analysis of household assets and house infrastructure. Life expectancy estimates for each wealth decile, sex, and social group are calculated using standard life table procedures. Estimates use sample weights. The shaded areas around the lines represent 95% confidence intervals calculated using a cluster-bootstrap procedure. Source: Annual Health Survey 2010-2011.

Figure 5: Demographic re-weighting: Unaccounted-for gaps in life expectancy (years) between OBC/high-caste Hindus and marginalized groups



Marginalized groups are re-weighted to reflect the same distribution of characteristics as among OBC/high-caste Hindus in three ways. 'Full difference' shows the full life expectancy gap between higher-caste Hindus and marginalized social groups (HC e_0 - Adivasi e_0, for instance). 'Unexplained by rural' shows the gap that remains after accounting for differences in rural residence. '+ economic status' shows the gap that remains after additionally accounting for differences in wealth index quintile intersected with land ownership. '+ environmental factors' shows the gap that remains after additionally accounting for differences in environmental exposures. For children under age five, this includes household solid fuel use intersected with four categories of PSU open defecation. For individuals age five and older, environmental factors only include household solid fuel use. The vertical lines around each estimate represent 95% confidence intervals calculated using a cluster-bootstrap procedure. Source: Annual Health Survey 2010-2011.

Appendix

A1 Appendix: Methods

A1.1 Estimating mortality rates and life expectancy

The AHS was designed to generate estimates of death rates and other demographic indicators at the district level, thus it has a large sample that is well-suited to estimating age-specific mortality rates. The household roster recorded the date of birth and sex of all members living in surveyed households on January 1, 2010. The survey also included a mortality roster, which recorded sex, age at death, and date of death, among other characteristics, of any usual members of surveyed households that died between January 1, 2007 and December 31, 2009. For individuals who were alive in January 2010, we estimate exact age using the date of birth provided in the household roster. For household members that died during the period, we use the exact age at death and date of death provided on the mortality roster. Using this information, we construct a dataset that records the number of person-years each individual contributed to each single-year age during the period January 2007 through December 2009, and also includes an indicator for whether or not the individual died at that age. The procedure we use to construct this dataset is discussed in greater detail in Appendix A2.4

With this dataset, we estimate age-specific mortality rates separately for each sex-by-social group using the following equation:

$$ {}_n m_x^{s,g} = \frac{\sum_{i=1}^{I} (died\ between\ age\ x\ to\ x+n_i) \times (weight_i)}{\sum_{i=1}^{I} (person\ years\ lived\ between\ age\ x\ to\ x+n_i) \times (weight_i)}. \tag{2} $$

In this equation, m refers to mortality rate, x to the interval's beginning age, and n to the length of the age interval. We use the abridged life table age ranges used by India's Sample Registration System (SRS).[1] i refers to individuals of sex s and social group g, and I to the total number of individuals of sex s and social group g covered in the survey. Observations are weighted by the sample weight provided in the survey to make the data representative of the nine AHS states. The reference period for deaths and person-years lived is January 2007 to December 2009.

We use standard life table procedures to estimate life expectancies for each social group, as well as each state-by-social group, and wealth decile-by-social group. To estimate life expectancies by additional characteristics, we modify Equation 2 to generate mortality rates for individuals defined by age group, sex, social group, and additional characteristics z, ${}_n m_x^{s,g,z}$. In our analyses, z represents states or wealth decile.

The number of person-years lived by those who died in age intervals x to $x + n$, or ${}_n a_x$, is calculated based on values from SRS 2007-2011 life tables for states. In practice, the use of different ${}_n a_x$, for instance, ${}_n a_x$ directly estimated from the AHS, changes life expectancy estimates very little (Preston, Heuveline and Guillot, 2000, pg. 47). The SRS does not produce life tables for different social groups. We therefore estimate group and sex-specific ${}_n a_x$ values for the overall sample of nine states by taking a weighted average of ${}_n a_x$ across

[1]The age intervals in years are as follows: zero to one, one to five, five to ten, ..., 80 to 85, and 85+.

states, where each state's weight is the fraction of deaths of individuals of the social group that occurred in the state.[2] Within states, for each of the social groups, we use state and sex-specific ${}_na_x$ from the SRS. We estimate wealth decile, group, and sex-specific ${}_na_x$ by taking a weighted average of ${}_na_x$ across states, where each state's weight is the fraction of deaths of individuals of a particular social group and wealth decile that occurred in the state.

A1.2 Demographic re-weighting

The re-weighting function, defined in Equation 1 of the main paper, is directly applied in the estimation of counterfactual age-specific mortality rates for marginalized groups. Counterfactual age-specific mortality rates are:

$$ {}_nm_x^{s,MG,CF} = \frac{\sum_{i=1}^{I}(died\ between\ age\ x\ to\ x+n_i) \times (weight_i) \times \psi^{MG}(c_i)}{\sum_{i=1}^{I}(person\ years\ lived\ between\ age\ x\ to\ x+n_i) \times (weight_i) \times \psi^{MG}(c_i)}, \tag{3} $$

where MG represents the marginalized group considered and HC represents OBC/high-caste Hindus. c_i represents a vector of observable characteristics for individual i that are correlated with life expectancy, including sex, age group, social group, and SES. ψ^{MG} is the re-weighting function. We use counterfactual age-specific mortality rates to estimate counterfactual life expectancies for each marginalized group.

Within each bin defined by sex, age group, and social group, we split the sample into additional bins based on exhibiting various combinations of SES. In the first scenario, the sample is split into two bins based on whether the individual lives in a rural area or not. The second scenario creates 20 bins by adding wealth status in five quintiles and a binary indicator for land ownership. The third scenario additionally includes environmental exposures. For children under age five, this includes household solid fuel use and four levels of PSU open defecation, generating 160 bins. For individuals age five and over, environmental exposures include only household solid fuel use, generating 40 bins.[3] The characteristics included in the re-weighting exercise are determined based on regression analysis, discussed in Appendix A1.3.

It is important to note that the unexplained gap in life expectancy should not be interpreted as the gap that can be attributed to discrimination or oppression for at least two reasons. First, differences in observed SES may arise from discriminatory processes. If this is the case, then the unexplained gap would underestimate the portion that can be attributed to discrimination. Second, the socioeconomic characteristics that we use to re-weight the sample of marginalized groups are necessarily limited to the characteristics measured in the survey. There may be other relevant characteristics that are not accounted for. If this is the

[2]The SRS did not produce life tables for the AHS states of Chhattisgarh, Jharkhand, or Uttarakhand. For Chhattisgarh, we use ${}_na_x$ values from Madhya Pradesh life tables, since Chhattisgarh was earlier a part of Madhya Pradesh. Similarly, we use ${}_na_x$ values from Bihar life tables for Jharkhand. Although Uttarakhand was formed from the state of Uttar Pradesh, the demography of Uttarakhand is very different, and so we use ${}_na_x$ values from India as a whole for this state.

[3]We exclude sanitation as an explanatory factor for adult mortality because sanitation is unlikely to have direct adult mortality consequences.

case, then the unexplained gap in life expectancy may overestimate the portion that can be attributed to discrimination.

A1.3 Description of linear regression

To motivate the re-weighting exercise, we present a regression analysis of the characteristics associated with mortality. We estimate these regressions separately by age, rather than in a pooled sample, because we hypothesize that the characteristics that affect child mortality may be different from the characteristics that affect adult mortality. Moreover, because the life table estimates mortality separately for different age groups, it is possible to use different characteristics for children and adults in the re-weighting exercise.

We estimate linear probability models predicting death separately for children under age five, and individuals age five and older. We cluster standard errors at the level of the primary sampling unit, and use survey weights. We estimate regressions of the following form:

$$\begin{aligned} died_{ihv} = & \boldsymbol{group_{hv}\beta} + \\ & \alpha rural_v + \boldsymbol{economic_{hv}\delta} + \boldsymbol{environmental_{hv}}\gamma + \\ & \boldsymbol{\theta_{ihv}} + \epsilon_{ihv}. \end{aligned} \tag{4}$$

where *died* is a dummy variable taking on values 0 or 1,000, indicating that the individual died between January 2007 and December 2009; *group* is a vector of dummy variables indicating social group with OBC/high-caste Hindus as the omitted category; *rural* is a dummy variable indicating rural residence; *economic* is a vector including wealth decile and a dummy variable indicating land ownership; and *environmental* is a vector including dummy variables for household open defecation and household solid fuel use, and continuous variables indicating the fraction of individuals in the PSU defecating in the open and the fraction eating food cooked on solid fuels. θ represents a vector of life table age group fixed effects. For regressions predicting child mortality, θ also includes fixed effects for potential exposure in months, which is the total number of months the child could have lived from January 2007 through December 2009. For children who were born before January 2007, potential exposure is 36 months. For children born after January 2007, potential exposure is the number of months the child could have lived before December 2009.

A2 Appendix: Data preparation

Annual Health Survey (AHS) data have been made available to the public by the Government of India.[4] Data from the household roster are available in "COMB" datasets, and data from the mortality roster are available in "MORT" datasets. We use data on usual residents of households surveyed in the first round of the AHS survey, which collected data in 2010-2011.

[4]Data are available from this website: https://nrhm-mis.nic.in/hmisreports/AHSReports.aspx.

A2.1 Social group variables

In order to ensure that there is no overlap in social group categories, we first classify households based on caste group, then by religion. Therefore, Muslims are individuals in Muslim households in the "other" caste group category. Similarly, higher-caste Hindus are individuals in Hindu households in the "other" caste group category. This classification is valid because it is rare for households in the data to be recorded as Muslim and Dalit, or Muslim and Adivasi.

Muslim Dalits and Muslim Adivasis are each 1.3% of the total Muslim sample. Dalit Muslims are 1% of the total Dalit sample and Adivasi Muslims are 1.9% of the total Adivasi sample. The households that do not fall in any of these four social groups are those in the "other" category that are neither Hindu nor Muslim. Individuals in these households represent less than one percent of the total sample of individuals.

Although OBCs are not as privileged in Indian society as high-caste Hindus, they were not considered untouchable, as Dalits were. As we note in the main text, we are unable to separate OBCs and high-caste Hindus. This makes our estimates for disparities conservative. The India Human Development Survey (IHDS) has noted the relative size of these groups (Desai et al., 2010). In India as a whole, OBCs constituted 36% of the population, and high-castes constituted 21% (pg. 224).

We are unable to assign 0.03% of observations to social groups because data are missing for *social group code*, which records caste, or *religion*. 93% of observations for which social group is unknown are from the mortality roster, and leaving them out based on missing data biases mortality estimates downwards. We therefore randomly assign these observations to social groups based on the existing distribution of the sample population across groups.

A2.2 Socioeconomic status variables

Following Filmer and Pritchett (2001), we construct a wealth index through principal component analysis of household ownership of assets (including radio, television, landline, mobile phone, washing machine, refrigerator, sewing machine, bicycle, motorcycle, car, tractor, water pump) and house infrastructure (number of rooms in the house, cooking location, and materials used to construct house). For some analyses we divide the wealth index into deciles, and for others, we divide the index into quintiles. We separately control for land ownership as another measure of wealth. For environmental exposures, we use indicators for household open defecation and use of solid fuels for cooking, as well as PSU-level fraction defecating in the open and cooking on solid fuels.

A2.3 Comparison of AHS mortality rates with SRS and NFHS

In addition to the AHS, we also use reported summary mortality data from India's Sample Registration System (SRS) for the years 2007 through 2009, and estimated age-specific mortality rates from India's Demographic and Health Survey 2015-2016, also called the National Family Health Survey (NFHS).[5] Across all nine states, the AHS visited over 20,000 primary

[5]Additional details on these three datasets, the deaths questions asked, and the estimation processes are available in Appendix Table A1.

sampling units (PSU)[6] and over four million households. The survey collected data on over 20 million individuals. In comparison, the SRS produced estimates of age-specific mortality rates and life expectancies in 2018 from a nationally representative sample of eight million individuals across most Indian states (Registrar General and Census Commissioner of India, 2020). Similarly, the National Family Health Survey (NFHS) collected data from approximately 600,000 households in 2015-2016 (International Institute for Population Sciences - IIPS/India and ICF., 2017) across all Indian states. The AHS is larger than these other national surveys, and was in fact one of the largest data collection exercises in the world.

The main purpose of referencing these alternative datasets is to show that AHS mortality rates match closely to those from the SRS and NFHS. The AHS is the primary data source for this paper because it contains data on social group, which the SRS does not, and is adequately large to study differences in mortality based on social group and SES, which the NFHS is not.

A2.4 Estimating person-years contributed to single-year ages

For individuals who were alive in January 2010, we estimate exact age using the date of birth provided in the household roster. For household members that died between January 2007 and December 2009, we estimate exact age at death and date of death using the data provided for each of these fields on the mortality roster.

Because the day of the month for both date of birth and date of death is missing for the majority of observations, we randomly assign a day of the month for all records. Some observations are also missing month of the year. Month of birth is missing for 45% of individuals alive on January 2010, and month of death is missing for 1% of individuals who died between January 2007 and December 2009. For observations for which month is missing, we randomly assign a month of the year.

Using data on day, month, and year of birth, we estimate the exact age of individuals who were alive on January 2010, and the number of person-years each individual contributed to single-year ages during the period January 2007 to December 2009. For individuals who died between January 2007 and December 2009, we use data from the fields on age at death, and data on day, month, and year of death, to estimate the number of person-years that these individuals contributed to single-year ages during the study period before they died.

A2.5 Identifying Primary Sampling Units (PSUs)

Primary sample units (PSUs) in the AHS are Census Enumeration Blocks (CEBs) in urban areas and villages in rural areas. Based on conversations with a statistician in the Government of India's Ministry of Statistics and Programme Implementation (MOSPI), we learned that for most records, PSU identifiers can be extracted from *fid* or *fidx* variables. Records either have a value under *fid* or a value under *fidx*. If the record has a value for *fid*, PSU is the three digit number starting at the fifth digit of the 12 or 13 digit number. If the record has a value for *fidx*, PSU is the last three digits of the 11 or 12 digit number.

[6] Sample units in the AHS are Census Enumeration Blocks (CEBs) in urban areas, and villages in rural areas.

For records that do not have a value for *fid* or *fidx*, PSU can be extracted using a third method. Because data were sorted by *state* and PSU before assigning the variable *psu id*, sorting the data by *state* and *psu id* can help in the process of backing out PSU for the remaining records. We sort the data by these two variables and assign records to PSUs based on the PSU of the surrounding records. If records that are above have a different PSU than records that are below an observation for which PSU is unknown, we randomly assign to the observation the PSU of one of the surrounding PSUs.

A3 Description of appendix figures and tables

Appendix Figure A1 shows the distribution of social groups by wealth deciles. We note that all deciles have representation from each of the four social groups. As expected, marginalized social groups are more likely to be in poorer deciles.

Appendix Figure A2 shows age-specific mortality rates for the four social groups. Muslims have lower or similar mortality rates compared to higher-caste Hindus until about age 50. After age 50, higher-caste Hindus have lower mortality rates. Dalits and Adivasis have higher mortality rates across the life course.

Appendix Figure A3 shows unadjusted mortality rates and predicted Gompertz mortality rates. Predicted mortality is similar to unadjusted mortality. Gaps between unadjusted and predicted mortality rates are slightly larger for marginalized social groups than for higher-caste Hindus. This is likely because of greater age-misreporting among marginalized social groups and suggests that the unadjusted life expectancy gaps reported in this study are conservative.

Appendix Table A2 provides a description of the survey sample. Across all nine states, the AHS visited over 20,000 primary sampling units (PSU) and over four million households. The survey collected data on over 20 million individuals. Some states have small populations of certain marginalized groups. For instance, Adivasis represent a small fraction of the population in Uttarakhand, Uttar Pradesh, and Bihar, and Muslims represent a small fraction of the population in Odisha and Chhattisgarh.

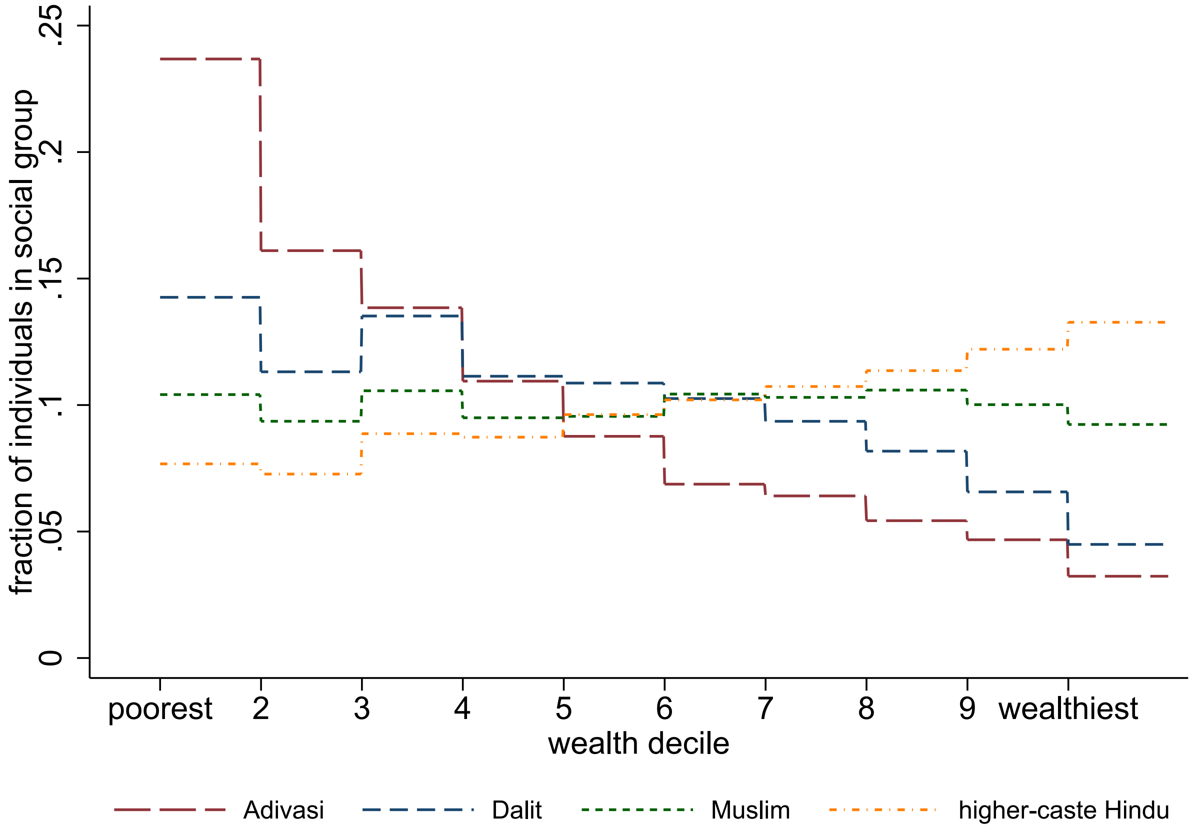
Appendix Table A3 shows summary statistics by social group. Panel A shows entries and exits in the AHS sample between January 2007 and December 2009. More than 19 million individuals were alive in January 2007. Over the three-year period, there were almost 1.4 million births and 446,085 deaths in sample households. More than 20 million individuals were alive at any point in this study period, contributing over 59 million person-years.

Panel B of Table A3 presents characteristics of sample individuals alive at any point between January 2007 and December 2009. Of this sample, 2.2% of individuals died in the period. Relative to the fraction of OBC/high-caste Hindu individuals in the sample that died, a higher fraction of Adivasis and Dalits in the sample died, and a lower fraction of Muslims in the sample died. 48.5% of the sample were female, and the average age was 23.9. The next set of characteristics in Panel B reflect SES. With the exceptions of Muslim advantage in urban residence, owning sewing machines, and sanitation, and Adivasi advantage in land ownership, this set of characteristics reflects that Adivasis, Dalits, and Muslims are disadvantaged across various socioeconomic indicators. Adivasis, Dalits, and Muslims own fewer assets, live in poorer quality homes, and are less likely to cook using clean fuels.

Appendix Table A4 shows results from linear regressions predicting the probability of death of children under age five. The regression results motivate the variables included in the decomposition. Column 4 includes all of the SES variables. Rural residence, wealth decile, land ownership, PSU and household open defecation, and household solid fuel use are predictive of child mortality. In order to maintain a manageable number of bins in the re-weighting exercise, we use PSU open defecation only, rather than both household and PSU open defecation variables, since we hypothesize that there are important externalities associated with open defecation.

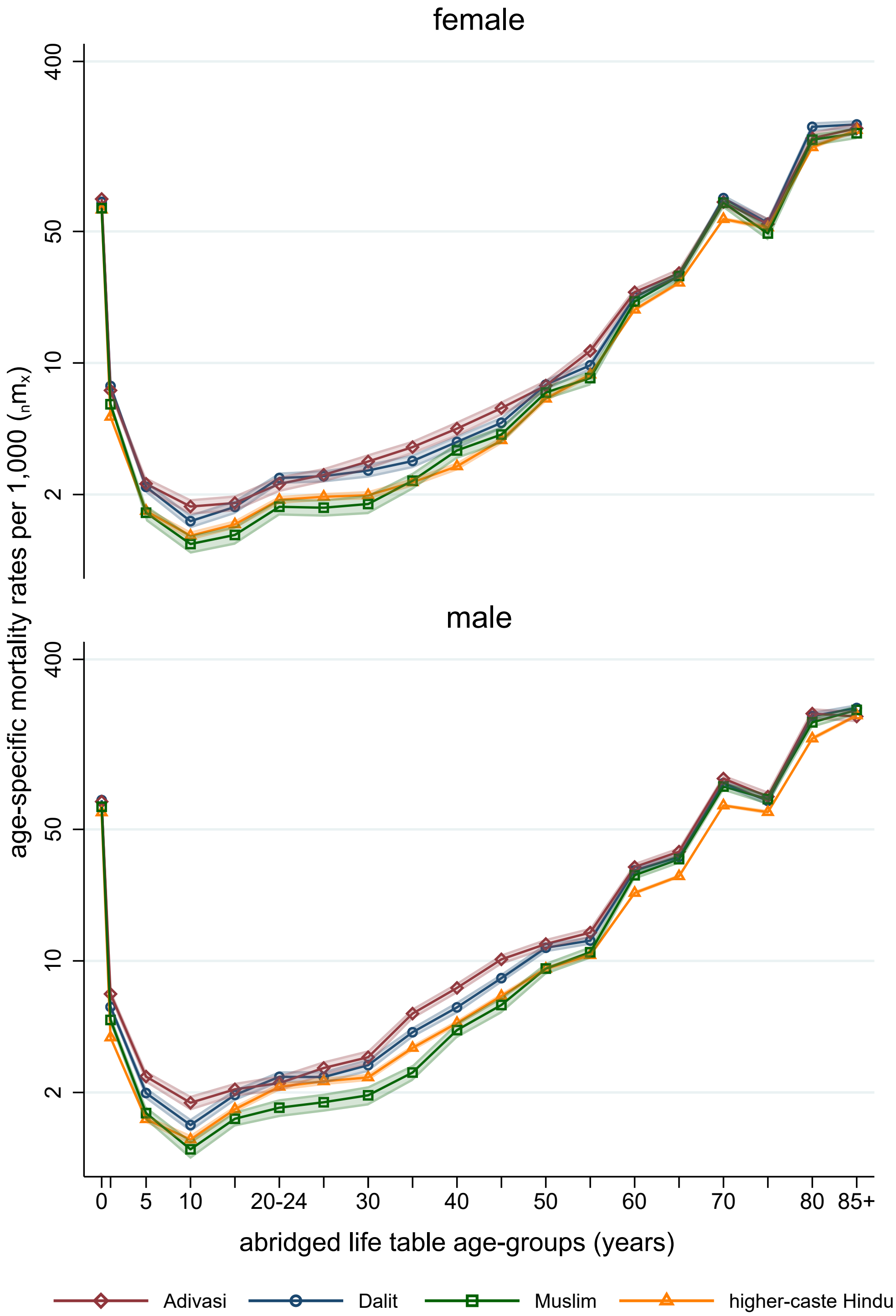
Appendix Table A5 shows results from linear regressions predicting the probability of death of individuals age five and older. The regression results motivate the variables included in the decomposition. Column 4 includes all of the SES variables. Rural residence, wealth decile, land ownership, and household solid fuel use are predictive of mortality.

Figure A1: Distribution of social groups across wealth deciles



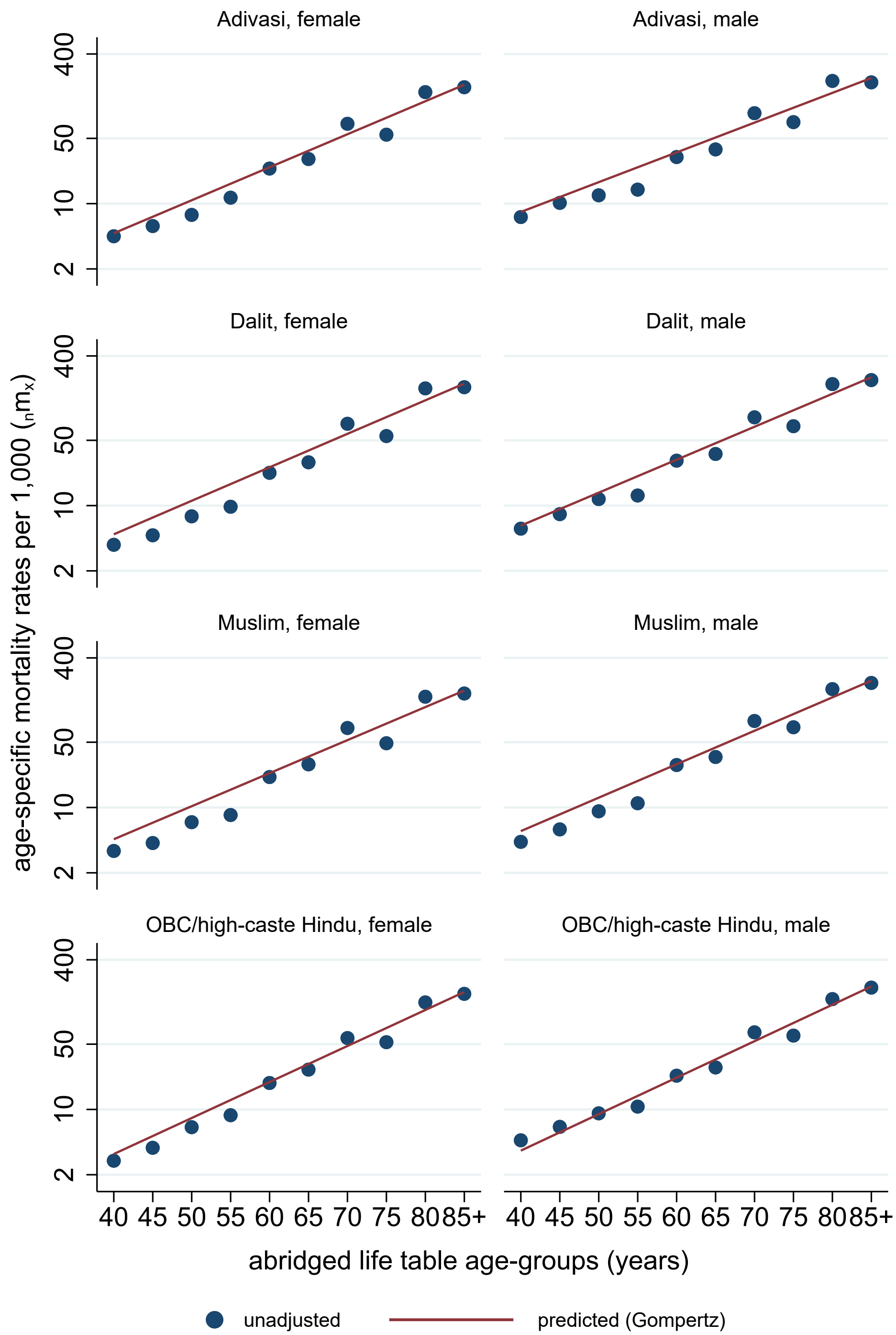
Wealth decile are deciles of a wealth index constructed using a principal component analysis of household assets and house infrastructure, including radio, television, landline, mobile phone, washing machine, refrigerator, sewing machine, bicycle, motorcycle, car, tractor, water pump, number of rooms in house, presence of a kitchen, and materials of house. Estimates use sample weights. Source: Annual Health Survey 2010-2011.

Figure A2: Age-specific mortality rates by sex and social group from AHS data



Age-specific mortality rates are based on deaths and person-years lived during the period Jan 2007 to Dec 2009, estimated from information provided in the mortality and household rosters. Estimates use sample weights. The shaded areas around the lines represent 95% confidence intervals calculated using a cluster-bootstrap procedure. Source: Annual Health Survey 2010-2011.

Figure A3: Actual and Gompertz mortality rates estimated from AHS data



Unadjusted age-specific mortality rates are based on deaths and person-years lived during the period Jan 2007 to Dec 2009, estimated from information provided in the mortality and household rosters. Estimates use sample weights. Gompertz mortality rates are estimated using a non-linear least squares model of the form: $m_x = \alpha \times e^{\beta x}$, where x represents age in five-year groups starting at age 40, with the final age interval as 85+ (Tai and Noymer, 2018). Source: Annual Health Survey 2010-2011.

Table A1: Data descriptions: AHS, SRS, and NFHS

Data source (1)	**Survey design** (2)	**Death questions used** (3)	**Sample size** (4)	**Reference period** (5)
AHS	sample survey: multi-stage stratified random sampling	household deaths question, instruction to enumerators: "For deaths occurred to the usual residents of the sample unit during January 1, 2007 to December 31, 2009. Still births not to be included." No further instruction beyond this.	4.1 million households in 9 states	Jan 2007 - Dec 2009
SRS	dual record: complete enumeration of births and deaths in a panel of villages, and semi-annual independent survey	we use published state-level age-specific mortality rates for years 2007, 2008, 2009	1.3 million households nationally	2007, 2008, 2009
NFHS	sample survey: multi-stage stratified random sampling	for individuals age 5 and older: household deaths question, "Did any usual member of this household die since Jan 2012?" for children under age 5: birth history recording information on all live births	0.6 million households nationally	Jan 2012 - Dec 2016

Mortality and life expectancy estimates generated by the SRS are used by both the UN-Inter-Agency Group for Child Mortality Estimation (UN-IGME) and the WHO Life Tables (World Health Organization, 2016; UN-IGME, 2018). Following Gupta and Sudharsanan (2020), in analyses that use data from the NFHS, we use the birth history to estimate child mortality and the mortality roster for estimating mortality at ages five and older. Sources: Annual Health Survey 2010-2011; Sample Registration System 2007, 2008, 2009; and National Family Health Survey 2015-2016.

Table A2: Description of Annual Health Survey sample

	Uttarakhand	**Rajasthan**	**Uttar Pradesh**	**Bihar**	**Assam**	**Jharkhand**	**Odisha**	**Chhattisgarh**	**Madhya Pradesh**	**All states**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Sample PSUs and households										
# of PSUs	2,501	1,842	3,927	2,356	1,784	2,109	2,365	1,255	2,557	20,696
# of households	367,177	352,744	847,675	594,834	383,413	378,919	457,725	273,609	494,616	4,150,712
PSU mean households	147	192	216	252	215	180	194	218	193	201
Panel B: Sample individuals alive at any time between Jan 2007 and Dec 2009										
Adivasi	0.048	0.145	0.017	0.022	0.129	0.250	0.214	0.306	0.209	0.102
Dalit	0.214	0.173	0.232	0.183	0.098	0.136	0.186	0.128	0.166	0.188
Muslim	0.125	0.094	0.190	0.168	0.313	0.155	0.020	0.017	0.067	0.144
OBC/high-caste Hindu	0.591	0.570	0.557	0.626	0.437	0.453	0.573	0.542	0.547	0.558
Other	0.022	0.018	0.005	0.001	0.022	0.006	0.008	0.007	0.010	0.008
n	1,652,464	1,826,710	4,662,156	3,155,938	1,772,163	1,960,716	1,981,435	1,246,948	2,353,231	20,611,761

Estimates use sample weights. PSU is primary sampling unit, which are Census Enumeration Blocks (CEBs) in urban areas, and villages in rural areas. Source: Annual Health Survey 2010-2011.

Table A3: Summary statistics by social group: socioeconomic disadvantage for marginalized social groups

	Adivasi (1)	**Dalit** (2)	**Muslim** (3)	**OBC/high-caste Hindu** (4)	**Other** (5)	**All** (6)
Panel A: Sample entries and exits between Jan 2007 and Dec 2009						
n (alive in Jan 2007)	2,321,293	3,410,534	2,423,716	10,916,863	165,691	19,238,097
births during period	175,871	266,052	196,687	726,854	9,644	1,375,108
deaths during period	59,775	85,842	50,382	246,425	3,661	446,085
n (alive at any time during period)	2,496,993	3,676,303	2,620,191	11,642,953	175,321	20,611,761
n person years	7,152,510	10,517,483	7,506,547	33,506,363	506,507	59,189,409
Panel B: Individuals alive at any time between Jan 2007 and Dec 2009						
died between Jan 2007 and Dec 2009	0.024	0.024	0.020	0.022	0.021	0.022
female	0.490	0.482	0.490	0.484	0.480	0.485
age (on Jan 2007)	23.269	22.884	21.357	24.914	27.643	23.872
household:						
in rural area	0.900	0.822	0.685	0.767	0.550	0.779
in bottom 40% of wealth index	0.673	0.541	0.439	0.359	0.228	0.437
owns land	0.633	0.461	0.336	0.545	0.336	0.510
has radio	0.132	0.155	0.182	0.224	0.291	0.196
has television	0.159	0.226	0.264	0.356	0.591	0.300
has computer	0.037	0.048	0.061	0.080	0.149	0.067
has phone	0.360	0.542	0.641	0.690	0.857	0.621
has washer	0.015	0.018	0.050	0.057	0.222	0.046
has refrigerator	0.036	0.051	0.110	0.129	0.401	0.103
has sewing machine	0.070	0.105	0.213	0.196	0.402	0.169
has bicycle	0.470	0.506	0.497	0.557	0.536	0.530
has motorcycle	0.088	0.093	0.137	0.203	0.385	0.162
has car	0.015	0.014	0.023	0.042	0.126	0.032
has tractor	0.012	0.011	0.015	0.032	0.084	0.024
has water pump	0.051	0.054	0.056	0.108	0.137	0.085
has >1 room in dwelling	0.544	0.467	0.486	0.624	0.755	0.568
has a kitchen	0.454	0.343	0.506	0.508	0.769	0.472
is kaccha	0.627	0.350	0.241	0.247	0.176	0.306
uses solid fuel for cooking	0.933	0.895	0.803	0.785	0.530	0.823
in PSU with >50% solid fuel use	0.940	0.900	0.815	0.822	0.604	0.847
defecates in the open	0.834	0.768	0.444	0.614	0.272	0.643
in PSU with >50% open defecation	0.851	0.780	0.482	0.694	0.333	0.697

Sample entries and exits are estimated from information provided in the mortality and household rosters. Wealth decile are deciles of a wealth index constructed using a principal component analysis of household assets and house infrastructure, including radio, television, landline, mobile phone, washing machine, refrigerator, sewing machine, bicycle, motorcycle, car, tractor, water pump, number of rooms in house, presence of a kitchen, and materials of house. Kaccha means households built using mud, straw, bamboo, and other temporary materials. Fraction open defecation and fraction solid fuel use in the PSU represent the PSU-level fraction of individuals defecating in the open and eating food cooked on solid fuels. Estimates use sample weights. Source: Annual Health Survey 2010-2011.

Table A4: Regressions predicting death, for individuals < age five on Jan 2007

Dependent variable:	Probability of death (per thousand) between Jan 2007 and Dec 2009			
	(1)	(2)	(3)	(4)
Social group (omitted category is OBC/high-caste Hindu)				
Dalit	6.463**	6.119**	4.721**	4.280**
	(0.451)	(0.444)	(0.462)	(0.469)
Adivasi	7.459**	6.341**	3.693**	3.426**
	(0.479)	(0.478)	(0.473)	(0.474)
Muslim	2.765**	3.734**	2.973**	3.961**
	(0.494)	(0.488)	(0.452)	(0.436)
Rural		10.87**	6.661**	1.848*
		(0.743)	(0.777)	(0.829)
Wealth decile			-1.452**	-1.008**
			(0.0530)	(0.0612)
Household owns land			1.831**	0.884**
			(0.323)	(0.325)
Fraction open defecation in PSU				2.681**
				(0.983)
Fraction solid fuel in PSU				1.057
				(1.317)
Household defecates in the open				4.403**
				(0.351)
Household uses solid fuel				4.170**
				(0.535)
Potential exposure in months	x	x	x	x
Age × sex fixed effects	x	x	x	x
N	3,709,562	3,709,562	3,709,562	3,709,562
Mean of dependent variable	32.398	32.398	32.398	32.398
Adjusted R-squared	0.0202	0.0208	0.0212	0.0214

Wealth decile are deciles of a wealth index constructed using a principal component analysis of household assets and house structure, including radio, television, landline, mobile phone, washing machine, refrigerator, sewing machine, bicycle, motorcycle, car, tractor, water pump, number of rooms in house, presence of a kitchen, and materials of house. Fraction open defecation and fraction solid fuel use in the PSU represent the PSU-level fraction of individuals defecating in the open and eating food cooked on solid fuels. Potential exposure in months is the total number of months the child could have lived from Jan 2007 through Dec 2009. For children who were born before Jan 2007, potential exposure is 36 months. For children born after Jan 2007, potential exposure is the number of months the child could have lived before Dec 2009. Age groups are abridged life table age-groups (years). Regressions use sample weights and cluster standard errors by PSU. Source: Annual Health Survey 2010-2011.

Table A5: Regressions predicting death, for individuals $\geq$ age five on Jan 2007

Dependent variable:	Probability of death (per thousand) between Jan 2007 and Dec 2009			
	(1)	(2)	(3)	(4)
Social group (omitted category is OBC/high-caste Hindu)				
Dalit	3.417**	3.363**	2.698**	2.596**
	(0.148)	(0.149)	(0.142)	(0.138)
Adivasi	4.416**	4.249**	3.345**	3.332**
	(0.237)	(0.235)	(0.237)	(0.235)
Muslim	0.689**	0.830**	0.350*	0.251
	(0.160)	(0.164)	(0.166)	(0.165)
Rural		1.293**	0.391	-0.506+
		(0.227)	(0.249)	(0.283)
Wealth decile			-0.502**	-0.413**
			(0.0244)	(0.0234)
Household owns land			-0.609**	-0.797**
			(0.123)	(0.120)
Fraction solid fuel in PSU				-0.117
				(0.403)
Household uses solid fuel				2.400**
				(0.159)
Age $\times$ sex fixed effects	x	x	x	x
N	16,886,515	16,886,515	16,886,515	16,886,515
Mean of dependent variable	19.428	19.428	19.428	19.428
Adjusted R-squared	0.0914	0.0915	0.0916	0.0916

Wealth decile are deciles of a wealth index constructed using a principal component analysis of household assets and house infrastructure, including radio, television, landline, mobile phone, washing machine, refrigerator, sewing machine, bicycle, motorcycle, car, tractor, water pump, number of rooms in house, presence of a kitchen, and materials of house. Fraction open defecation and fraction solid fuel use in the PSU represent the PSU-level fraction of individuals defecating in the open and eating food cooked on solid fuels. Age groups are abridged life table age-groups (years). Regressions use sample weights and cluster standard errors by PSU. Source: Annual Health Survey 2010-2011.