Large and unequal life expectancy declines associated with the COVID-19 pandemic in India in 2020

Aashish Gupta, Payal Hathi, Murad Banaji, Prankur Gupta, Ridhi Kashyap, Vipul Paikra, Kanika Sharma, Anmol Somanchi, Nikkil Sudharsanan, and Sangita Vyas*

May 29, 2023

*Aashish Gupta (aashish.gupta@sociology.ox.ac.uk) is at the University of Oxford. Payal Hathi (phathi@berkeley.edu) is at the University of California, Berkeley. Murad Banaji (murad.banaji@maths.ox.ac.uk) is at the University of Oxford. Prankur Gupta (prankur@utexas.edu) is at the University of Texas at Austin, and is an affiliate of the Population Wellbeing Initiative at UT Austin. Ridhi Kashyap (ridhi.kashyap@sociology.ox.ac.uk) is at the University of Oxford. Vipul Paikra (vipul@riceinstitute.org) is at the research institute for compassionate economics. Kanika Sharma (kanika.sharma@emory.edu) is at Emory University. Anmol Somanchi (anmol.smnch@psestudent.eu) is at the Paris School of Economics. Nikkil Sudharsanan (nikkil.sudharsanan@tum.de) is at the Technical University Munich. Sangita Vyas (sangita.vyas@hunter.cuny.edu) is at CUNY Hunter College, and is an affiliate of the CUNY Institute for Demographic Research and the Population Wellbeing Initiative at UT Austin. For helpful comments and suggestions, we are grateful to Jose-Manuel Aburto, Lisa Berkman, Diane Coffey, Jennifer Dowd, Jean Dréze, Dennis Feehan, Victor Manuel Garcia Guerrero, Apoorva Jadhav, David Levine, Anup Malani, Dean Spears, and S.V. Subramanian. This research was supported by the Leverhulme Trust (Grant RC-2018-003) for the Leverhulme Centre for Demographic Science; by a National Institute of Child Health and Human Development Training grant at the University of California, Berkeley, T32HD007275; by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Individual Fellowship (agreement No. 101027598); and the Economic and Social Research Council through the Connecting Generations Centre grant (ref ES/W002116/1). Additional support was provided by the research institute for compassionate economics (r.i.c.e.).

Abstract

The toll of the COVID-19 pandemic in low- and middle-income countries (LMICs) is uncertain. We are the first to use high-quality empirical data from India, which has a large population and where pandemic surveillance was particularly poor, to examine changes in period life expectancy and estimate excess deaths during the pandemic. We analyze data from the households interviewed in 2021 in India's fifth Demographic and Health Survey, a subsample representative of about one-quarter of India's population. In this subsample, life expectancy at birth declined by 2.6 years between 2019 and 2020, a reduction that is larger than the loss in life expectancy observed in any high-income country (HIC) in the same period. Mortality was 17.0% higher in the pandemic months of 2020 compared to 2019. Applied nationally, this level of excess mortality implies 1.18 million excess deaths in 2020. Compared to HICs, mortality increases in younger ages in India contributed more to the decrease in life expectancy than older ages. Furthermore, the pandemic exacerbated gender and social inequalities. In contrast to global patterns, females in India experienced larger life expectancy losses than males. Among social groups, relative to a life expectancy loss of 1.5 years for high caste Hindus, who are privileged in Indian society, Muslims lost 5.9 years, Scheduled Tribes lost 4.4 years, and Scheduled Castes lost 2.6 years. These findings uncover large and unequal mortality impacts during the pandemic in the world's most populous country.

Keywords: COVID-19, excess mortality, life expectancy, disparities, India

1 Introduction

The COVID-19 pandemic generated a global mortality shock, resulting in large losses in period life expectancy (hereafter, life expectancy) worldwide.[1] In high-income countries (HICs), high-quality pandemic surveillance and vital registration systems documented substantial life expectancy declines (Schöley et al., 2022; Aburto et al., 2022c) and increased disparities across socioeconomic groups (Aburto et al., 2022c; Andrasfay and Goldman, 2021; Billingsley et al., 2022; Case and Deaton, 2023; Goldman and Andrasfay, 2022; Luck et al., 2022). However, much remains unknown about the scale and social gradient of COVID-19 mortality in low- and middle-income countries (LMICs), where limited resources imply poor emergency health response, pandemic surveillance, and data quality (Helleringer and Queiroz, 2022; Heuveline, 2022; Ho, 2021; Nepomuceno et al., 2020).

This study is the first to empirically estimate changes in life expectancy at birth by sex and social group between 2019 and 2020 in India, where according to the WHO (Knutson et al., 2022), one-third of all excess deaths are estimated to have occurred. We also estimate monthly excess mortality in 2020 relative to baseline. To do this, we use high-quality empirical data on mortality and socioeconomic characteristics from India's fifth Demographic and Health Survey (DHS), known as the National Family Health Survey-5 (NFHS-5). This exceptionally large dataset helps to address major gaps in knowledge about pandemic mortality in India that stem in part from incomplete administrative data and low-quality survey data (Banaji et al., 2022; Guilmoto, 2022; Drèze and Somanchi, 2021; Rukmini, 2020).

We first establish the credibility of NFHS mortality rates by comparing them to rates from official sources. Pre-pandemic mortality estimates from the NFHS closely match life tables from the United Nations and the Government of India's nationally-representative mortality surveillance system (Gupta, 2021; Gupta and Sudharsanan, 2022). Data collection for the fifth round of the NFHS was carried out between 2019 and 2021. We use the subsample of households interviewed in 2021 to study mortality in 2020 relative to prior years. This subsample includes households from 14 states and union territories. It is representative of about one-fourth of India's population and is similar to the full sample in terms of demographic and socioecomomic characteristics. Unless otherwise noted, all analyses are for the NFHS-5 2021 subsample. In the remainder of this text, we refer to it as the "subsample".

[1] Period life expectancy is a summary measure of mortality in a period that enables comparisons of the mortality impacts of the pandemic across populations of different sizes and age structures.

We find a 2.6-year decline in life expectancy at birth between 2019 and 2020 in the subsample. This decline is larger than the decline in modeled life expectancy estimates in India (Heuveline, 2022; United Nations, 2022). It is also more than the loss in life expectancy in any HIC in the same period (Schöley et al., 2022; Aburto et al., 2022b). In HICs, life expectancy declines were primarily driven by mortality increases in older age groups above 60 (Aburto et al., 2022b). In contrast, mortality increases below age 60 contributed more to the decline in life expectancy in India, even as mortality increased across the life course.

Our findings demonstrate that the toll of the pandemic was experienced unevenly within India. Whereas in most countries, losses to life expectancy were greater for males than females (Aburto et al., 2022b; Geldsetzer et al., 2022), we document a loss in life expectancy among females that is one year more than for males. Larger declines among females relative to males are partly explained by greater mortality increases in younger ages among females, particularly during childhood and the early reproductive years.

We also find greater life expectancy declines among disadvantaged caste and religious groups relative to privileged social groups in the subsample. Indian society is one of the most stratified in the world. Scheduled Castes (SCs), Scheduled Tribes (STs), and Muslims face social marginalization based on caste, indigenous identity, and religion, respectively (Ambedkar, 1945; Xaxa, 1999; Sachar Committee, Government of India, 2006). Relative to a decline in life expectancy of 1.5 years for high caste Hindus, who are privileged in Indian society, the loss for Muslims was 5.9 years, for STs was 4.4 years, and for SCs was 2.6 years. Before the pandemic began, each of these three groups faced large disadvantages in life expectancy at birth relative to high caste Hindus (Vyas et al., 2022; Gupta and Sudharsanan, 2022). The pandemic exacerbated these disparities. These declines are comparable or larger in absolute magnitude to those experienced by Native Americans, Blacks, and Hispanics in the United States in 2020 (Goldman and Andrasfay, 2022; Andrasfay and Goldman, 2021).

We find that mortality in the subsample was 17% higher in the pandemic months of 2020 relative to 2019, and was particularly elevated in the last four months of 2020. These estimates are robust to alternative baselines. While likely smaller than the 2021 mortality surge (Banaji and Gupta, 2022), our results show that if the rest of India also experienced an increase in 2020 mortality similar to the subsample, it would imply 1.18 million excess deaths in 2020 nationally. Relative to other estimates during the same period, our extrapolated estimate for all-India excess deaths is about eight times the official number of COVID-19 deaths in India (covid19india.org, 2023), 1.5 times the WHO's extrapolated estimate of excess deaths in India (Knutson et al., 2022), and more than 2.5 times the estimated excess deaths in the United States in 2020 (Ahmad et al., 2021).

Our estimates using high-quality NFHS data fill an important gap in scientific understanding of COVID-19 mortality in India and globally in 2020. Administrative data from India's Civil Registration System, which have been used for many existing estimates of excess mortality in India (Banaji and Gupta, 2022; Jha et al., 2022; Anand et al., 2021), including the WHO estimates (Knutson et al., 2022; Msemburi et al., 2023), do not capture all births and deaths, are unavailable for many states, and were disrupted by India's severe lockdown in 2020. Although informative for highlighting weaknesses of administrative data, existing surveys that have been used for prior research on India's pandemic mortality are not representative (Somanchi, 2021) and do not produce reliable estimates of baseline mortality (Malani and Ramachandran, 2022; Anand et al., 2021).

Because India is the largest country in the world, understanding the global toll of the pandemic relies on accurately estimating pandemic mortality in India. Our findings uncover large and unequal effects of the pandemic in India, and show that disadvantages can be exacerbated in times of a mortality crisis. Methodologically, our analysis demonstrates the value of empirical approaches using high-quality data to measure routine and crisis mortality. In particular, our analysis underscores that large-scale sample surveys which ask questions on recent deaths of household members are valuable for mortality surveillance in data sparse settings. These approaches may reveal empirical patterns missed by modeling approaches or non-representative and low-quality data sources.

2 Data and methods

2.1 The main sample: NFHS-5 subsample of households interviewed in 2021

This study uses data from the fifth round of the nationally-representative National Family Health Survey (NFHS-5), which is India's Demographic and Health Survey. In particular, we use the subsample of households interviewed in 2021 to study mortality in 2020 relative to prior years. Data collection for the NFHS-5 was scheduled over two phases. Phase-1 states were interviewed in 2019 and the first two months of 2020. Interviews in Phase-2 states began in late 2019, but were disrupted by the severe lockdown in March 2020. Households interviewed in Phase 1 states and in Phase 2 states before the lockdown in March 2020 are *not* part of the main analysis sample because they were interviewed before the pandemic began. Data collection resumed in Phase-2 states in October 2020 and continued until May 2021 when the surge in mortality due to the delta wave prevented further interviews. As Figure SI-1 shows, the Phase-2 households interviewed in 2021 comprise the main analysis sample for this study.[2] This subsample is representative of 23.2% of the Indian population.

Figure SI-2 shows the proportion of interviews conducted in each calendar month for each of the 36 states and union territories where NFHS-5 was fielded, and Figure SI-3 displays a map of primary sampling units (PSUs) that are and are not part of the main analysis sample. Out of 36 states, interviews were conducted in early 2021 in 14 states and union territories: Punjab, Chandigarh, Uttarakhand, Haryana, Delhi, Rajasthan, Uttar Pradesh, Arunachal Pradesh, Jharkhand, Odisha, Chhattisgarh, Madhya Pradesh, Tamil Nadu, and Puducherry.

Tables SI-1 and SI-2 compare the 2021 subsample to the full NFHS-5 sample. The subsample is similar to the full sample in terms of distribution by urban and rural residence, sex, and the proportion of the population that identifies as Scheduled Caste. There are minor differences between the subsample and the full sample in the proportion of the population that is Hindu, Muslim, Scheduled Tribe, and Other Backward Class. The subsample is also slightly younger than the full sample, with a greater proportion of the population that is below age five.

We estimate mortality for 2018, 2019, and 2020 all from the same households interviewed in 2021 using retrospective questions on mortality. Results are therefore not biased by comparing mortality pre- and post-pandemic between groups that differ on characteristics. Although the 2021 subsample is geographically clustered and the spread of the pandemic may have varied spatially, Table SI-3 shows that disease spread at the end of 2020, as measured in the third national serosurvey (Murhekar et al., 2021), was similar in states visited by the NFHS-5 in 2021 compared to states visited earlier. We also show that our results are not driven by any single state, and are robust to excluding one state at a time in hypothetical sub-subsamples (subsamples of our NFHS-5 2021 subsample) in Figure SI-4.

2.2 Other data

We use several other data sources for comparative purposes and computations:

- Age-specific mortality rates from the Sample Registration System (SRS), which is the Government of India's nationally-representative mortality surveillance system

[2] We do not include the Phase 2 state households interviewed in the last quarter of 2020 because we estimate mortality for these months.

(Registrar General and Census Commissioner of India, 2013). The SRS collects data on all vital events in sample villages. This sample surveillance system is distinct from India's Civil Registration System (CRS), which is meant to record all vital events nationally but is still incomplete because many births and deaths are not registered (Saikia et al., 2023; Banaji et al., 2022).

- WHO estimates of excess mortality from Msemburi et al. (2023). Estimates for India are constructed based on data from India's CRS, which in addition to being incomplete, are only available for a portion of the pandemic period for 17 states and union territories. Additionally, the CRS was disrupted by India's severe lockdown in 2020.
- Recorded COVID-19 cases and COVID-19 deaths, which were reported by labs and hospitals to state agencies, and compiled by covid19india.org (2023).
- Seroprevalence from Murhekar et al. (2021), a nationally-representative serosurvey carried out in December 2020 to January 2021.
- Estimated age-specific mortality rates, deaths, and age distribution of the Indian population from United Nations World Population Prospects (United Nations, 2022).

2.3 Estimating age-specific mortality rates and life expectancy at birth

We follow the approach described in Gupta (2021) and Gupta and Sudharsanan (2022) to estimate age-specific mortality rates from the NFHS. We use standard demographic approaches (Moultrie et al., 2013; Gupta, 2022) to estimate mortality under age five from the birth history module of NFHS-5. For mortality at age five and above, we use questions on deaths in the household since January 2017 to estimate deaths and person-years lived by those who did not survive. In addition to the age and sex of each deceased household member, information on the month and year of death was also collected. For those who were alive at the time of the interview, we use the household roster, which contains the age of each individual in the household, to estimate person-years lived. We estimate age-specific mortality rates by creating empirical life-lines at the calendar-month level for those who died and for those who were alive at the time of the survey.

Figure 1 shows that age-specific mortality rates for 2018-2019 computed from the full NFHS-5 sample (not the subsample) using this approach are similar to those in the same period from the United Nations and the Sample Registration System (SRS), which is the

Government of India's official mortality monitoring system.[3] 95% confidence intervals are shown as the shaded area around the NFHS estimates.[4]

We estimate age-, sex-, and social-group specific mortality rates for 2019 and 2020 separately. Using standard approaches, we construct period life tables to calculate life expectancy at birth in 2019 and 2020 separately. The average number of person-years lived by those who died in an age-interval, or ${}_na_x$, is borrowed from 2015-2019 official SRS life tables.

We first compare life expectancy at birth in 2019 to 2020 for the full subsample, and separately for females and males. To explore disparities in life expectancy changes, we estimate life expectancy at birth in 2019 and 2020 by social group, following the same social categories as Desai et al. (2010) and Gupta and Sudharsanan (2022). We compare high caste Hindus, who are relatively privileged in Indian society, to four marginalized social groups: Scheduled Castes (SCs), Scheduled Tribes (STs), Muslims, and Other Backward Classes (OBCs).

To assess the importance of changes in mortality in different age groups to the overall change in life expectancy at birth between 2019 and 2020, we apply Arriaga's decomposition (Arriaga, 1984). We estimate the contribution of differences in mortality between ages 0-19, 20-39, 40-59, 60-79, and 80+ to the change in life expectancy between 2019 and 2020 for females and males separately.

2.4 Estimating excess mortality

To study excess mortality in 2020 relative to prior years, we estimate crude death rates for each calendar month between January 2018 and December 2020. To do this, we estimate age-specific deaths and person-months in each calendar month using the NFHS-5 2021 subsample. We annualize these estimates of monthly death rates in order to enable comparisons across periods that vary in length. Annualized monthly death rates can be interpreted as expected mortality in a year if the mortality in the month under consideration was observed for an entire year. Crude death rates have also been age-standardized using the 2020 population age-distribution, estimated by the United Nations World Pop-

[3]For some age groups, particularly in young adulthood, NFHS-5 mortality rates are higher than those in the SRS. In addition to the fact that SRS microdata are not made publicly available, SRS reports do not describe the methods used for constructing age-specific mortality rates. In the past, NFHS and SRS age-specific mortality rates were more similar, with both showing a mortality hump in young adulthood (Gupta and Sudharsanan, 2022). Without more information on SRS mortality estimation procedures, it is not possible to reject the possibility that the SRS actually underestimates mortality in young adulthood in 2018-2019. Franz et al. (2022) show that the SRS underestimates maternal mortality relative to the NFHS.

[4]SRS and UN do not provide confidence intervals for age-specific mortality rates.

ulation Prospects (United Nations, 2022).

Figure SI-5 shows crude death rates calculated using this approach for each calendar month in 2018, 2019, and 2020. For each month in 2020, we use these crude death rates to calculate the excess mortality P-score, which is the percentage increase in mortality in a particular month relative to that same month in the baseline period. As baseline mortality, we consider mortality observed in 2019, and average mortality observed in 2018 and 2019.

Recall bias is a concern using retrospective mortality data from household surveys. Figure SI-6 shows mortality rates relative to the month of interview using data from households interviewed in 2019. The figure shows that the level of mortality reported in months more proximate to the interview is not higher than the level of mortality in months further in the past. Therefore, excess mortality in 2020 relative to 2019, or higher mortality in the latter part of 2020, cannot be explained by recall bias in the NFHS-5 2021 subsample.

All analyses use survey weights. To calculate 95% confidence intervals, we use a cluster-bootstrap approach that replicates the multi-stage sampling structure of NFHS-5 and accounts for clustering of observations within primary sampling units (PSUs) (Cameron and Miller, 2015).

3 Results

3.1 Life expectancy declines between 2019 and 2020 by age and sex

Figure 2 shows estimates of life expectancy at birth in 2019 and 2020 for females, males, and the combined population, and Figure SI-7 shows estimates of the change between the two periods. The vertical lines represent 95% confidence intervals. Between 2019 and 2020, overall life expectancy at birth declined by 2.6 years [95% CI: 1.8 to 3.4] in the NFHS-5 2021 subsample.[5] This drop is greater than the decline in life expectancy observed in the same period in any HIC (Schöley et al., 2022). It is also larger than the reduction between 2019 and 2020 for India modeled by the United Nations (United Nations, 2022) and others (Heuveline, 2022). Discrepancies across studies may in part be due to variation in the quality of the underlying data.[6]

[5] Figure SI-8 shows that this decline was similar in rural and urban areas.

[6] It is unlikely that different mortality patterns in non-subsample regions explain the difference between modeled estimates and NFHS-5 subsample estimates. Life expectancy at birth in non-subsample regions would have had to decline by as little as 0.1 to 0.4 years in order for NFHS-5 subsample estimates to be consistent with modeled estimates.

This overall decline masks substantial heterogeneity across gender. Increases in mortality over this period narrowed the life expectancy advantage observed among females in the subsample. Relative to a decline of 2.2 years [95% CI: 1.1 to 3.2] observed among males, the decline for females was one year larger at 3.2 years [95% CI: 1.9 to 4.4].[7] [8] These patterns are in stark contrast to the global pattern of a greater increase in mortality during the COVID-19 pandemic for males compared to females (Aburto et al., 2022a; Schöley et al., 2022; Geldsetzer et al., 2022).

Greater declines for females relative to males in India likely reflect gender inequality.[9] Prior research has documented that Indian households spend less on health care for females relative to males (Saikia and Bora, 2016; Dupas and Jain, 2021), a pattern which likely worsened during the pandemic (Jain and Dupas, 2022; Thejesh et al., 2021). Gender inequality is also suggested by the under-representation of females in India's official COVID-19 case data (Rukmini, 2020), despite similar levels of seroprevalence relative to males in sample surveys (Murhekar et al., 2021).

Figure 3 shows the absolute contribution, in years, of increases in mortality in different age groups to declines in life expectancy at birth between 2019 and 2020. The panel on the left shows results for females and the panel on the right shows results for males. 95% confidence intervals are indicated below each estimate. For both females and males, large mortality increases in younger age groups under age 60 contributed substantially to the reduction in life expectancy at birth. For females, mortality increases in the 0-19 and 60-79 ages contributed the most, approximately one year each [95% CIs: -1.8 to -0.2 and -1.5 to -0.4, respectively], to the decline. For males, mortality increases in the 40-59 ages contributed the most, 0.8 years [95% CI: -1.3 to -0.3], to the decline. These findings contrast sharply with age-specific contributions to changes in life expectancy in HICs, where most of the life expectancy declines in 2020 were due to mortality increases in the older age ranges above age 60, and especially age 80 and older (Aburto et al., 2022a).

[7] The 95% confidence interval for the difference in decline between females and males is -0.7 years to 2.7 years. Although we cannot rule out that the greater decline for females relative to males is due to sampling error, in the vast majority of countries, male life expectancy declined by more than female life expectancy. Such a large female disadvantage in the impact of the pandemic has not been documented in any country.

[8] Limited available sex-specific civil registration data on deaths from urban municipalities in Gujarat (Acosta et al., 2022) and from the south Indian city of Chennai (Lewnard et al., 2022) show that mortality increases among females and males were similar in 2020 compared to 2019. However, civil registration data are more likely to miss female deaths than male deaths (Saikia et al., 2023).

[9] Gender inequality in India has long been shown to reduce the female advantage that we would expect in the absence of discrimination against girls and women (Kashyap, 2019; Anderson and Ray, 2012). Consistent with gender inequality as an explanation, we observed larger losses among females compared to males among all social groups except for STs, among whom gender disparities are lower (Xaxa, 1999; Maharatna, 2000).

The relatively large contribution of mortality increases in the age range 0-19, particularly among females, may reflect mortality from causes other than COVID-19.[10] Prior research has identified pandemic disruptions to public health services such as childhood immunizations (Summan et al., 2023), tuberculosis treatment (Pai et al., 2022), and hospital births (Kumari et al., 2020). These disruptions might have been partly responsible for an increase in all-cause mortality in 2020. The increases in mortality below age 80 may also reflect a younger age-profile of COVID-19 mortality in India, as hypothesized by Nepomuceno et al. (2020) and Levin et al. (2022).

3.2 Greater life expectancy declines among marginalized social groups

India is a highly unequal society. For each social group group separately, Figure 4 shows estimates of life expectancy at birth in 2019 and 2020, and Figure SI-7 shows estimates of the change between the two years. Panel A shows estimates for the combined population, Panel B for females, and Panel C for males. Estimates for SCs are shown in the first column, STs in the second column, Muslims in the third column, OBCs in the fourth column, and high caste Hindus in the last column. Vertical lines represent 95% confidence intervals.

Marginalized social groups experienced the largest reductions in life expectancy at birth, which exacerbated already existing social inequalities. Relative to a 1.5 year [95% CI: -0.4 to 3.4] life expectancy loss among high-caste Hindus, the loss for Muslims was 5.9 years [95% CI: 3.1 to 8.8], for STs was 4.4 years [95% CI: 2.3 to 6.5], and for SCs was 2.6 years [95% CI: 0.8 to 4.4]. These declines are similar or greater in absolute magnitude than declines experienced by Blacks, Hispanics, and Native Americans in the United States during the pandemic (Aburto et al., 2022c; Andrasfay and Goldman, 2022; Goldman and Andrasfay, 2022). The decline for OBCs of 1.3 years [95% CI: -0.1 to 2.6] was similar in magnitude to that among high caste Hindus. In 2020, Muslim life expectancy was the lowest across the five social groups, a result of the fact that Muslims observed the greatest declines in life expectancy at birth between 2019 and 2020. This decline is consistent with the further marginalization of Muslims in 2020 (Yasir, 2020). [11]

[10] Figure SI-9, which shows age-specific contributions to life expectancy declines by abridged life table age groups, reveals that the declines for females were particularly large during childhood, the early reproductive years, and above age 60.

[11] Life expectancy in 2019 and 2020 by social group and urban or rural residence is displayed in Figure SI-10. It reveals that high caste Hindus in urban areas experienced the smallest decline in life expectancy (0.5 years), whereas Muslims in urban areas experienced the largest decline (8.4 years). For all marginalized groups, losses are apparent in both rural and urban areas, but the sample size is not large enough to identify where declines were larger.

3.3 Excess mortality in 2020

Figure 5 displays monthly estimates of 2020 excess mortality from the NFHS-5 2021 subsample, and compares these results to WHO estimates of excess mortality (Msemburi et al., 2023), and official records of case counts and deaths (covid19india.org, 2023). We compute excess mortality using two different baseline mortality estimates: the same calendar month in 2019, and the average over the same calendar month in 2018 and 2019. Vertical lines represent 95% confidence intervals.

Mortality was elevated from September to December 2020, relative to the same months in the previous years (Panel A, Figure 5). On average between April and December 2020, mortality was 17.0% [95% CI: 10.6 to 23.3] higher than in 2019. Excess mortality was higher for females than males. Table SI-4 estimates excess deaths nationally based on these findings. If excess mortality from the NFHS-5 2021 subsample were observed nationally, it would imply 1.18 million [95% CI: 0.74 to 1.63] excess deaths in 2020.

We verify that these results are robust in several ways (Nepomuceno et al., 2022; Shkolnikov et al., 2022). First, using mean mortality in 2018 and 2019 as an alternative baseline, the 2020 excess mortality P-score is 16.8% [95% CI: 11.3 to 22.2] in the pandemic months. Second, to ensure that these results are not driven by the particular states that are in the subsample, we calculate excess mortality P-scores by constructing hypothetical sub-subsamples (subsamples of our NFHS-5 2021 subsample) which exclude one state at a time (Msemburi et al., 2023; Banaji and Gupta, 2022). Figure SI-4 shows that excess mortality is similar across hypothetical subsamples.

Panel A of Figure 5 also compares NFHS-5 subsample excess mortality to WHO estimates, which use incomplete Civil Registration System data from 17 states and union territories, and extrapolate to the rest of the country. Both sources show elevated mortality from September through December 2020. However, NFHS-5 estimates do not display negative excess mortality in March, April, and May 2020, which the WHO estimates suggest. At least some of the reduction in mortality in the WHO data was likely due to a disruption of vital registration during the strict lockdown during these months (Banaji et al., 2022). This would not have affected mortality estimates from NFHS-5 because they are based on survey data.

Panels B and C display monthly COVID-19 cases and deaths reported by laboratories and hospitals to state agencies, and compiled by covid19india.org (2023). These show a peak in cases and deaths in September 2020 and a subsequent decline. This pattern is inconsistent with the NFHS-5 subsample and WHO estimates, which both show elevated mortality from September through December 2020. One reason for the discrepancy across data sources could arise from an increase in mortality from causes other than COVID-19 in the latter part of the year, which would show up in NFHS-5 and WHO estimates, but not pandemic surveillance. Another reason could be biases in pandemic surveillance (Banaji, 2021). Figure SI-5, which shows NFHS-5 estimates of annualized age-standardized crude death rates by month for urban and rural areas separately, suggests that official surveillance may have more closely tracked COVID-19 mortality in urban areas, and missed mortality dynamics in rural areas.

4 Discussion

This is the first study to use high-quality empirical data to understand the scale, distribution, and disparities in excess mortality during the first year of the pandemic in India, where a substantial fraction of global pandemic deaths are estimated to have occurred, but where impacts have been uncertain. In our subsample, which represents one-fourth of India's population, we find a reduction in life expectancy at birth of 2.6 years between 2019 and 2020, larger than reductions documented in any HIC. This reduction was substantial even relative to trends in India: overall life expectancy at birth in 2020 was equivalent to all-India levels over a decade earlier (Registrar General and Census Commissioner of India, 2013). From a comparative perspective, India's decline is similar to or larger than declines seen in the same period in other large LMICs including Brazil (Castro et al., 2021; Fernandes et al., 2023), Russia (Aburto et al., 2022a), and Mexico (García-Guerrero and Beltrán-Sánchez, 2021). Our estimates imply 1.18 million excess deaths in India nationally in 2020, which is about one-third of the excess deaths estimated by the WHO for the rest of the world (Msemburi et al., 2023).

The increase in mortality between 2019 and 2020 was heterogeneous across age, sex, and social group. Although mortality increased across all ages in India, relative to HICs, increases in mortality in younger age groups (under age 60) contributed more to life expectancy declines. In contrast to global patterns, we find that life expectancy declined by one full year more for females than for males. Like in the United States, where the pandemic increased existing gaps in life expectancy, such as those by race, ethnicity, and education (Case and Deaton, 2023; Andrasfay and Goldman, 2022; Aburto et al., 2022c), gaps in India between privileged and marginalized groups also increased. Relative to high caste Hindus, the gap in life expectancy at birth for Scheduled Castes increased from 4.8 years in 2019 to 5.9 years in 2020, for Scheduled Tribes it increased from 2.2 to 5.1 years, and for Muslims it increased from 2.2 to 6.6 years.

The NFHS-5 is a rich data source that uncovers patterns of mortality that could not be observed in other data sources. India's Civil Registration System (CRS), which records vital events, is incomplete even in normal times with lower coverage among females and marginalized groups (Saikia et al., 2023). It was also disrupted during India's severe lockdown in March through May of 2020 (Banaji and Gupta, 2022). These factors may explain why the WHO estimates, which are based on CRS data from 17 states and union territories, are biased downwards. Recorded case and death counts in India are also biased because they miss most infections and COVID-19 deaths (Murhekar et al., 2021; Banaji and Gupta, 2022; Zimmermann et al., 2021). They are also more likely to capture males than females (Rukmini, 2020; KC and Moradhvaj, 2023) due to gender inequality in access to healthcare (Saikia and Bora, 2016). We additionally find evidence that recorded case and death counts more closely tracked urban than rural excess mortality. Moreover, these data miss mortality patterns from causes other than COVID-19. These findings highlight the inherent risks in inferring mortality patterns from recorded COVID-19 case and death data. Finally, existing survey-based estimates of excess mortality are based on biased samples which underrepresent women, young children, rural areas, and the poor (Somanchi, 2021).

Our findings have important implications for further research. For example, more clarity is needed to understand why females in India fared worse than males in terms of life expectancy losses and excess mortality, why the age profile of excess mortality is younger in India than in other countries, and why Muslims suffered such high losses to life expectancy at birth relative to other social groups. Age and gender patterns showing that females and younger age groups in India have been disproportionately impacted by the pandemic suggest that increases in mortality were not solely caused by direct COVID-19 mortality. Additional data are needed to understand direct versus indirect mortality impacts of the COVID-19 pandemic in 2020, and beyond.

Another important remaining gap in our understanding relates to changes in mortality in regions that were not represented in the NFHS-5 2021 subsample. Although our 2021 subsample is similar to the full NFHS-5 sample on socioeceonomic and demographic characteristics, 2021 interviews are geographically clustered in 14 states. To the extent that the spread and mortality impact of the pandemic varied by geography, our results may or may not provide a full understanding of changes in mortality at the national level in 2020 compared to 2019. However, evidence suggests that changes in mortality in this subsample might not have been so different from regions that are not in this subsample. India's third national serosurvey (Murhekar et al., 2021) shows that disease spread was similar in the subsample states compared to states that were not in the subsample. In addition, our estimates of excess mortality are not driven by any one state, and are similar in hypothetical sub-subsamples (subsamples of the NFHS-5 2021 subsample), which exclude one state at a time (see Figure SI-4).

Methodologically, our study shows the potential for accurately estimating mortality, even in short time intervals, using retrospective mortality information collected in a large-scale sample survey in a relatively poor context. These approaches may be helpful in routine mortality surveillance and understanding other mortality crises using in-person or phone-surveys (Adjiwanou et al., 2020). The extent to which retrospective questions on deaths of household members estimate mortality well in other contexts is a topic for further research. In particular, biases arising from recall errors, non-coverage of households in which all members died, and household dissolution may be more important in other contexts (Vyas et al., 2022; Gupta, 2021; Timaeus, 1991; Hill, 1991).

From a policy perspective, it is clear that the pandemic exacerbated longstanding inequalities in population health, particularly along dimensions of caste, religion, indigenous identity, rural or urban residence, age, and sex. In showing that pandemics can exacerbate inequalities rather than level them (Mamelund and Dimka, 2021; Klein, 1973), our findings reinforce the relevance of perspectives that emphasize social conditions as a fundamental cause of health and mortality (Link and Phelan, 1995). Although pandemic mortality in India in 2020 did not receive the same attention as the 2021 surge due to the Delta variant, our results show large and unequal mortality increases even early in the pandemic. Overall, the findings suggest that a greater focus on disadvantaged groups such as females, marginalized populations, and rural areas is important in understanding and responding to future mortality crises.

References

Aburto, José Manuel, Jonas Schöley, Ilya Kashnitsky, and Ridhi Kashyap (2022a) "Life expectancy declines in Russia during the COVID-19 pandemic in 2020," *International Journal of Epidemiology*, 51 (5), 1695–1697.

Aburto, José Manuel, Jonas Schöley, Ilya Kashnitsky et al. (2022b) "Quantifying impacts of the COVID-19 pandemic through life-expectancy losses: a population-level study of 29 countries," *International journal of epidemiology*, 51 (1), 63–74.

Aburto, José Manuel, Andrea M Tilstra, Ginevra Floridi, and Jennifer Beam Dowd (2022c) "Significant impacts of the COVID-19 pandemic on race/ethnic differences in US mortality," *Proceedings of the National Academy of Sciences*, 119 (35), e2205813119.

Acosta, Rolando J, Biraj Patnaik, Caroline Buckee, Mathew V Kiang, Rafael A Irizarry, Satchit Balsari, and Ayesha Mahmud (2022) "All-cause excess mortality across 90 municipalities in Gujarat, India, during the COVID-19 pandemic (March 2020-April 2021)," *PLOS Global Public Health*, 2 (8), e0000824.

Adjiwanou, Vissého, Nurul Alam, Leontine Alkema et al. (2020) "Measuring excess mortality during the COVID-19 pandemic in low-and lower-middle income countries: the need for mobile phone surveys."

Ahmad, Farida B, Jodi A Cisewski, Arialdi Miniño, and Robert N Anderson (2021) "Provisional mortality data—united states, 2020," *Morbidity and Mortality Weekly Report*, 70 (14), 519.

Ambedkar, Bhimrao Ramji (1945) *Annihilation of caste with a reply to Mahatma Gandhi*.

Anand, Abhishek, Justin Sandefur, and Arvind Subramanian (2021) "Three new estimates of India's all-cause excess mortality during the COVID-19 pandemic."

Anderson, Siwan and Debraj Ray (2012) "The age distribution of missing women in India," *Economic and Political Weekly*, 87–95.

Andrasfay, Theresa and Noreen Goldman (2021) "Reductions in 2020 US life expectancy due to COVID-19 and the disproportionate impact on the Black and Latino populations," *Proceedings of the National Academy of Sciences*, 118 (5), e2014746118.

——— (2022) "Reductions in US life expectancy during the COVID-19 pandemic by race and ethnicity: Is 2021 a repetition of 2020?" *Plos one*, 17 (8), e0272973.

Arriaga, Eduardo E (1984) "Measuring and explaining the change in life expectancies," *Demography*, 21 (1), 83–96.

Banaji, Murad (2021) "What lessons do India's Covid-19 serosurveys hold?" *Scroll.in*, 06 February, https://scroll.in/article/986097/what-lessons-do-indias-covid-19-serosurveys-hold.

Banaji, Murad and Aashish Gupta (2022) "Estimates of pandemic excess mortality in India based on civil registration data," *PLOS Global Public Health*, 2 (12), e0000803.

Banaji, Murad, Aashish Gupta, and Vipul Paikra (2022) "Mortality and Death Registration in India: The many holes in the data," *The India Forum*, https://www.theindiaforum.in/article/mortality-and-death-registration-india.

Billingsley, Sunnee, Maria Brandén, Siddartha Aradhya, Sven Drefahl, Gunnar Andersson, and Eleonora Mussino (2022) "COVID-19 mortality across occupations and secondary risks for elderly individuals in the household: a population register-based study," *Scandinavian journal of work, environment & health*, 48 (1), 52.

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

Case, Anne and Angus Deaton (2023) "Accounting for the Widening Mortality Gap Between Adult Americans with and without a BA."

Castro, Marcia C, Susie Gurzenda, Cassio M Turra, Sun Kim, Theresa Andrasfay, and Noreen Goldman (2021) "Reduction in life expectancy in Brazil after COVID-19," *Nature Medicine*, 27 (9), 1629–1635.

covid19india.org (2023) "COVID-19 India - covid19india.org," Accessed April 2023.

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

Drèze, J. P. and A. Somanchi (2021) "Not having the have-nots," *Economic Times*, 21 June.

Dupas, Pascaline and Radhika Jain (2021) "Women left behind: gender disparities in utilization of government health insurance in India,"Technical report, National Bureau of Economic Research.

Fernandes, Fernando, Cássio M Turra, Giovanny VA França, and Marcia C Castro (2023) "Mortality by cause of death in Brazil: effects of the COVID-19 pandemic and contribution to changes in life expectancy at birth," *medRxiv*, 2023–02.

Franz, Nathan, Aashish Gupta, Dean Spears, and Diane Coffey (2022) "Uncertainty about maternal mortality in India: New, higher estimates from the National Family Health Survey-4."

García-Guerrero, Víctor M and Hiram Beltrán-Sánchez (2021) "Heterogeneity in excess mortality and its impact on loss of life expectancy due to COVID-19: evidence from Mexico," *Canadian studies in population*, 48 (2-3), 165–200.

Geldsetzer, Pascal, Trasias Mukama, Nadine Kamel Jawad, Tim Riffe, Angela Rogers, and Nikkil Sudharsanan (2022) "Sex differences in the mortality rate for coronavirus disease 2019 compared to other causes of death: an analysis of population-wide data from 63 countries," *European Journal of Epidemiology*, 37 (8), 797–806.

Goldman, Noreen and Theresa Andrasfay (2022) "Life expectancy loss among Native Americans during the COVID-19 pandemic," *Demographic Research*, 47, 233–246.

Guilmoto, Christophe Z (2022) "An alternative estimation of the death toll of the Covid-19 pandemic in India," *Plos one*, 17 (2), e0263187.

Gupta, Aashish (2021) *Health, Environment, and Inequality in India* Ph.D. dissertation, University of Pennsylvania.

——— (2022) "Seasonal variation in infant mortality in India," *Population Studies*, 1–18.

Gupta, Aashish and Nikkil Sudharsanan (2022) "Large and Persistent Life Expectancy Disparities between India's Social Groups," *Population and Development Review*.

Helleringer, Stephane and Bernardo Lanza Queiroz (2022) "Commentary: Measuring excess mortality due to the COVID-19 pandemic: progress and persistent challenges," *International journal of epidemiology*, 51 (1), 85–87.

Heuveline, Patrick (2022) "Global and National Declines in Life Expectancy: An End-of-2021 Assessment," *Population and Development Review*, 48 (1), 31–50.

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

Ho, Jessica Y (2021) "What Demographers Need—and What the World Needs from Demographers—in Response to Covid-19."

Jain, Radhika and Pascaline Dupas (2022) "The effects of India's COVID-19 lockdown on critical non-COVID health care and outcomes: Evidence from dialysis patients," *Social Science & Medicine*, 296, 114762.

Jha, Prabhat, Yashwant Deshmukh, Chinmay Tumbe et al. (2022) "COVID mortality in India: National survey data and health facility deaths," *Science*, 375 (6581), 667–671.

Kashyap, Ridhi (2019) "Is prenatal sex selection associated with lower female child mortality?" *Population studies*, 73 (1), 57–78.

KC, Samir and Moradhvaj (2023) "Impact of the COVID-19 pandemic on the age-sex pattern of COVID-19 deaths in India," *Asian Population Studies*, 1–20.

Klein, Ira (1973) "Death in India, 1871–1921," *The Journal of Asian Studies*, 32 (4), 639–659.

Knutson, Victoria, Serge Aleshin-Guendel, Ariel Karlinsky, William Msemburi, and Jon Wakefield (2022) "Estimating Global and Country-Specific Excess Mortality During the COVID-19 Pandemic," *arXiv preprint arXiv:2205.09081*.

Kumari, Vimla, Kalpana Mehta, and Rahul Choudhary (2020) "COVID-19 outbreak and decreased hospitalisation of pregnant women in labour," *The Lancet Global Health*, 8 (9), e1116–e1117.

Levin, Andrew T, Nana Owusu-Boaitey, Sierra Pugh et al. (2022) "Assessing the burden of

COVID-19 in developing countries: systematic review, meta-analysis and public policy implications," *BMJ Global Health*, 7 (5), e008477.

Lewnard, Joseph A, Ayesha Mahmud, Tejas Narayan, Brian Wahl, TS Selvavinayagam, Ramanan Laxminarayan et al. (2022) "All-cause mortality during the COVID-19 pandemic in Chennai, India: an observational study," *The Lancet Infectious Diseases*, 22 (4), 463–472.

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

Luck, Anneliese N, Samuel H Preston, Irma T Elo, and Andrew C Stokes (2022) "The unequal burden of the Covid-19 pandemic: Capturing racial/ethnic disparities in US cause-specific mortality," *SSM-Population Health*, 17, 101012.

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

Malani, Anup and Sabareesh Ramachandran (2022) "Using household rosters from survey data to estimate all-cause excess death rates during the COVID pandemic in India," *Journal of Development Economics*, 159, 102988.

Mamelund, Svenn-Erik and Jessica Dimka (2021) "Not the great equalizers: Covid-19, 1918–20 influenza, and the need for a paradigm shift in pandemic preparedness," *Population Studies*, 75 (sup1), 179–199.

Moultrie, Tom A, Rob E Dorrington, Allan G Hill, Kenneth Hill, Ian M Timæus, and Basia Zaba (2013) *Tools for demographic estimation*: International Union for the Scientific Study of Population.

Msemburi, William, Ariel Karlinsky, Victoria Knutson, Serge Aleshin-Guendel, Somnath Chatterji, and Jon Wakefield (2023) "The WHO estimates of excess mortality associated with the COVID-19 pandemic," *Nature*, 613 (7942), 130–137.

Murhekar, Manoj V, Tarun Bhatnagar, Jeromie Wesley Vivian Thangaraj et al. (2021) "SARS-CoV-2 seroprevalence among the general population and healthcare workers in India, December 2020–January 2021," *International Journal of Infectious Diseases*, 108, 145–155.

National Commission on Population (2019) "Population projections for India and States 2011-2036," *Report of the Technical Group on Population Projections Constituted by the National Commission on Population*.

Nepomuceno, Marília R, Enrique Acosta, Diego Alburez-Gutierrez, José Manuel Aburto, Alain Gagnon, and Cássio M Turra (2020) "Besides population age structure, health and other demographic factors can contribute to understanding the COVID-19 burden," *Proceedings of the National Academy of Sciences*, 117 (25), 13881–13883.

Nepomuceno, Marília R, Ilya Klimkin, Dmitri A Jdanov, Ainhoa Alustiza-Galarza, and Vladimir M Shkolnikov (2022) "Sensitivity Analysis of Excess Mortality due to the COVID-19 Pandemic," *Population and Development Review*, 48 (2), 279–302.

Pai, Madhukar, Tereza Kasaeva, and Soumya Swaminathan (2022) "Covid-19's devastating effect on tuberculosis care—A path to recovery," *New England Journal of Medicine*, 386 (16), 1490–1493.

Registrar General and Census Commissioner of India (2013) *Compendium of India's Fertility and Mortality Indicators ,1971 - 2013*: Government of India, https://censusindia.gov.in/vital_statistics/Compendium/Srs_data.html.

Rukmini (2020) "India's covid-19 count could be missing many women," *LiveMint*, https://www.livemint.com/news/india/india-s-covid-19-count-could-be-missing-many-women-11598670961443.html.

Sachar Committee, 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 Jayanta Kumar Bora (2016) "Gender difference in health-care expenditure: evidence from India human development survey," *PloS one*, 11 (7), e0158332.

Saikia, Nandita, Krishna Kumar, and Bhaswati Das (2023) "Death registration coverage 2019–2021, India," *Bulletin of the World Health Organization*, 101 (2), 102.

Schöley, Jonas, José Manuel Aburto, Ilya Kashnitsky, Maxi S Kniffka, Luyin Zhang, Hannaliis Jaadla, Jennifer B Dowd, and Ridhi Kashyap (2022) "Life expectancy changes since COVID-19," *Nature human behaviour*, 6 (12), 1649–1659.

Shkolnikov, Vladimir M, Ilya Klimkin, Martin McKee et al. (2022) "What should be the baseline when calculating excess mortality? New approaches suggest that we have underestimated the impact of the COVID-19 pandemic and previous winter peaks," *SSM-Population Health*, 18, 101118.

Somanchi, Anmol (2021) "Missing the poor, big time: A critical assessment of the consumer pyramids household survey."

Summan, Amit, Arindam Nandi, Anita Shet, and Ramanan Laxminarayan (2023) "The effect of the COVID-19 pandemic on routine childhood immunization coverage and timeliness in India: Retrospective analysis of the National Family Health Survey of 2019–2021 data," *The Lancet Regional Health-Southeast Asia*, 8, 100099.

Thejesh, GN, Krushna Ranaware, Kanika Sharma, and Aman (2021) "Of denial and data deaths due to India's covid-19 national lockdown," *Economic and Political Weekly*, 56 (23), 37–41.

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

United Nations (2022) "World population prospects 2022," *Department of Economic and Social Affairs: Population Division*.

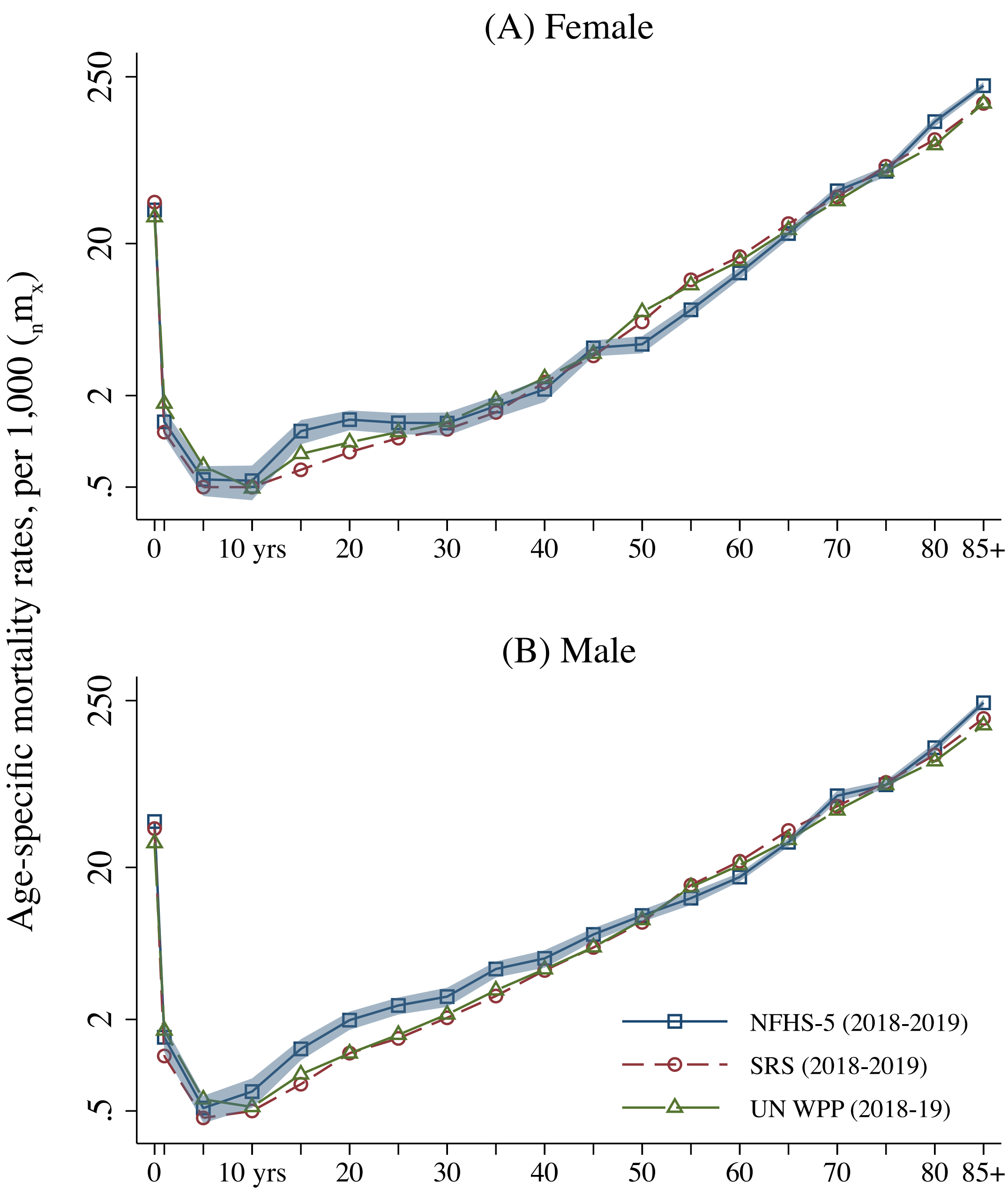
Vyas, Sangita, Payal Hathi, and Aashish Gupta (2022) "Social disadvantage, economic inequality, and life expectancy in nine Indian states," *Proceedings of the National Academy of Sciences*, 119 (10), e2109226119.

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

Yasir, Sameer (2020) "India is scapegoating Muslims for the spread of the Coronavirus," *Foreign Policy*, 22.

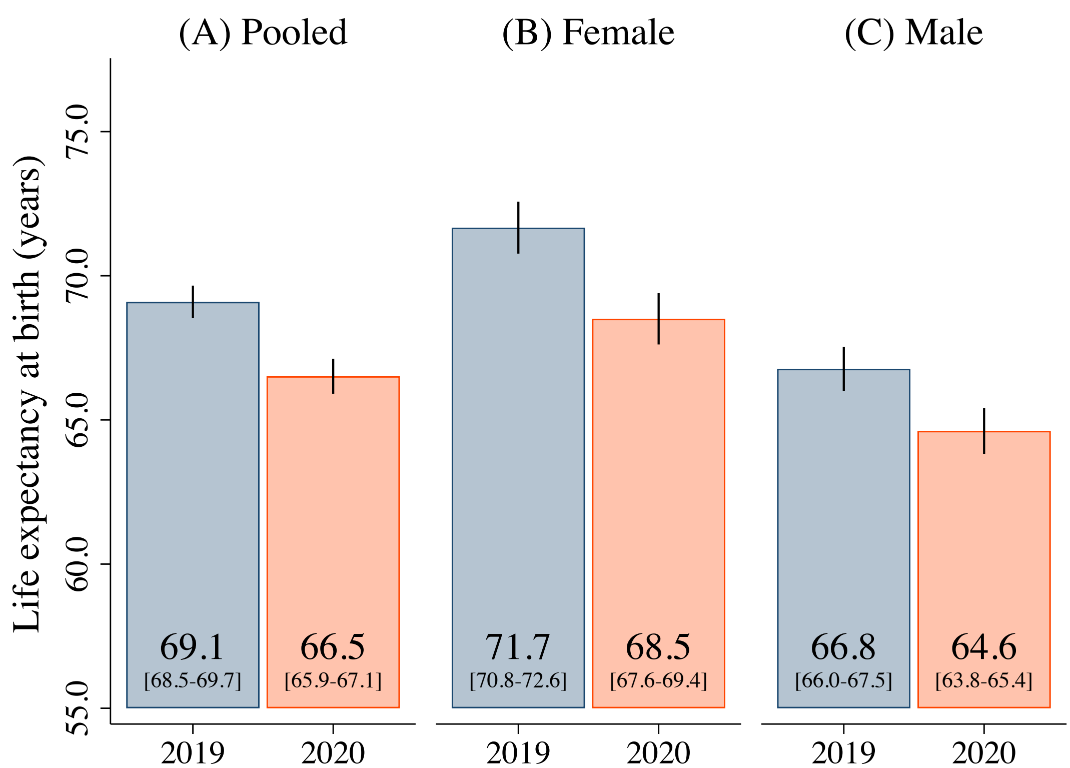
Zimmermann, Lauren V, Maxwell Salvatore, Giridhara R Babu, and Bhramar Mukherjee (2021) "Estimating COVID-19–related mortality in India: an epidemiological challenge with insufficient data," *American Journal of Public Health*, 111 (S2), S59–S62.

Figure 1: Similar age-specific mortality rates (${}_{n}m_{x}$), from National Family Health Survey-5, Sample Registration System and United Nations, 2018-2019



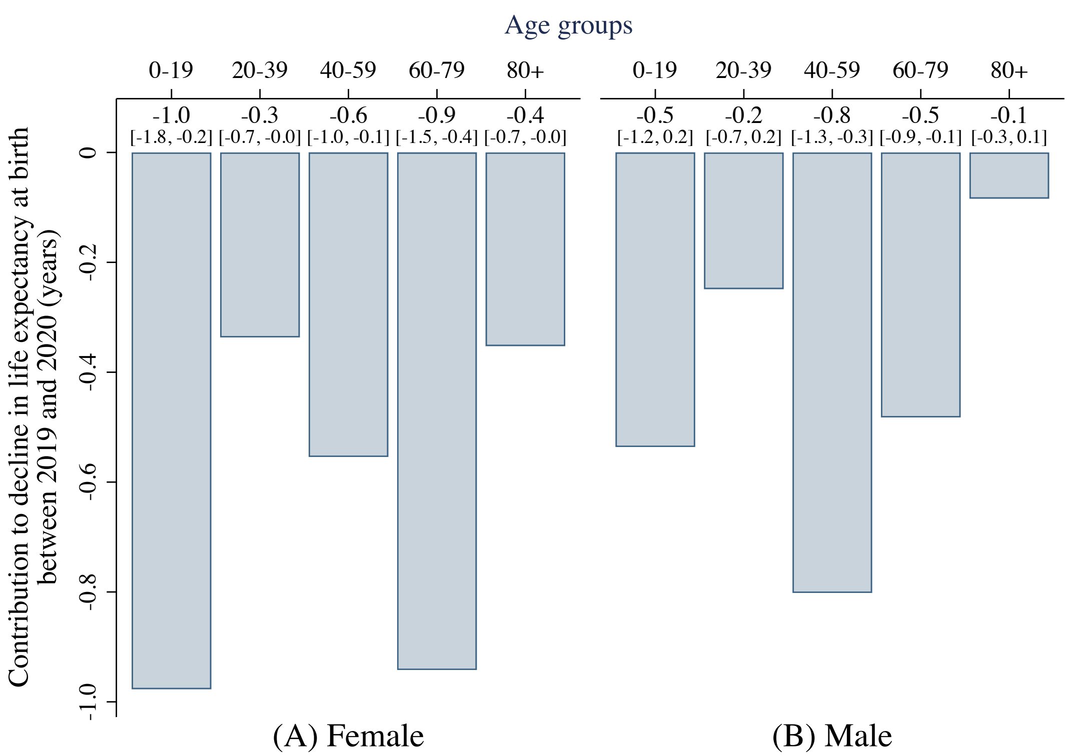
Note: Age-specific mortality rates are shown for (A) females and (B) males, separately, from three different sources. NFHS estimates are for the full NFHS-5 sample (not the subsample). For the NFHS-5, mortality rates are estimated based on the procedure described in the Methods section, using the full sample. NFHS-5 estimates use sample weights. SRS estimates are mean age-specific mortality rates published in SRS annual reports for 2018 and 2019. 95% confidence intervals for NFHS-5 estimates are shown as the shaded area around estimates and are calculated using a cluster-bootstrap approach. Confidence intervals for SRS are not shown because SRS microdata are not publicly available and the reports do not include clustered standard errors. Confidence intervals are not provided in UN data. Sources: NFHS-5, SRS 2018 and 2019, and UN WPP 2022.

Figure 2: Declines in period life expectancy between 2019 and 2020 are large and patterned by gender



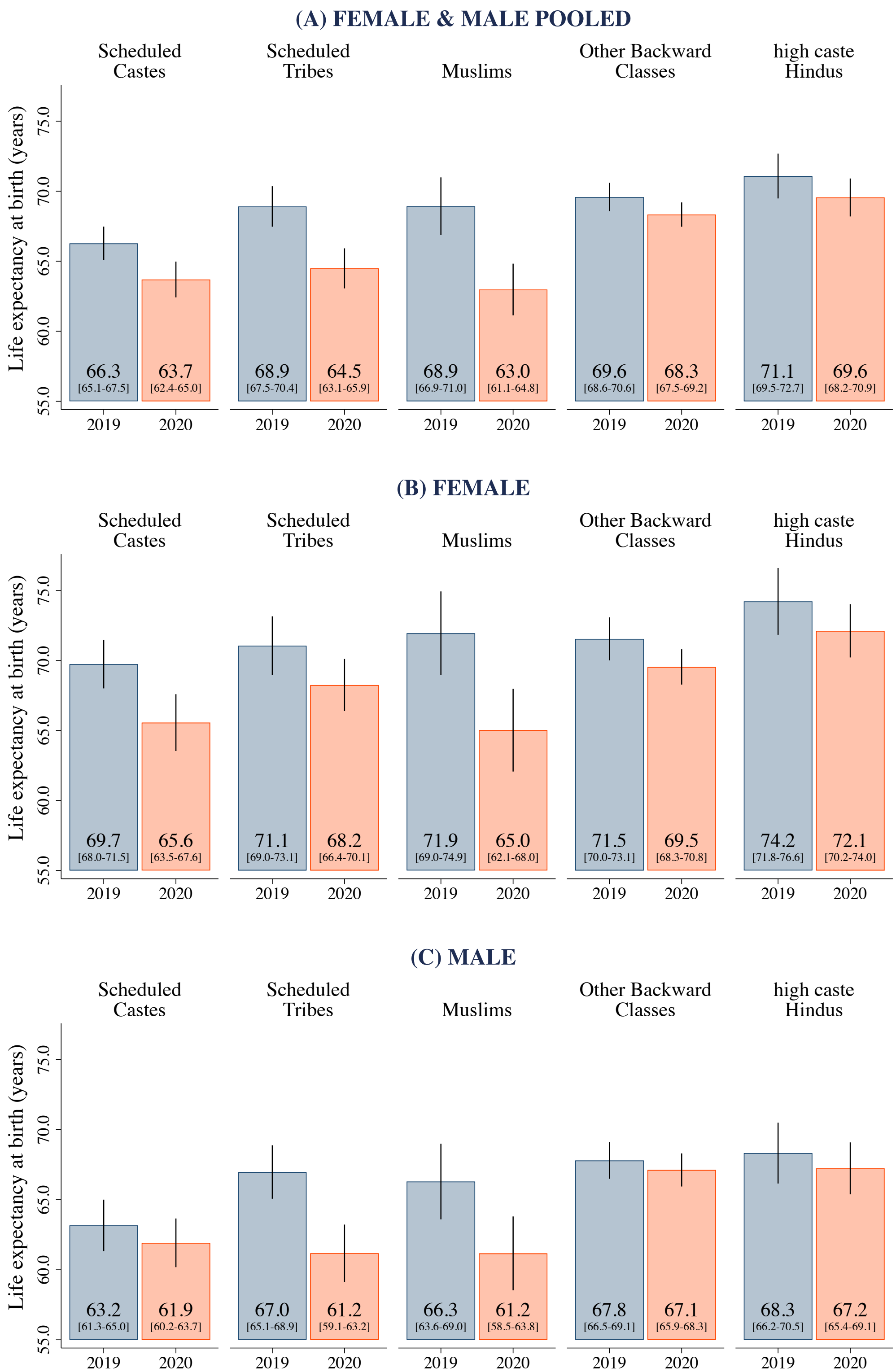
Note: The figure shows life expectancy at birth in 2019 and 2020 for the (A) combined female and male population, (B) females, and (C) males, separately. Life expectancy is calculated based on standard life table procedures. Estimates are for the NFHS-5 2021 subsample and use sample weights. The vertical lines around each estimate represent 95% confidence intervals calculated using a cluster-bootstrap approach. See Figure SI-7 for estimates of the decline in life expectancy between the two years. Source: NFHS-5.

Figure 3: Increases in mortality in younger age-groups contributed more than older age-groups to declines in period life expectancy at birth between 2019 and 2020



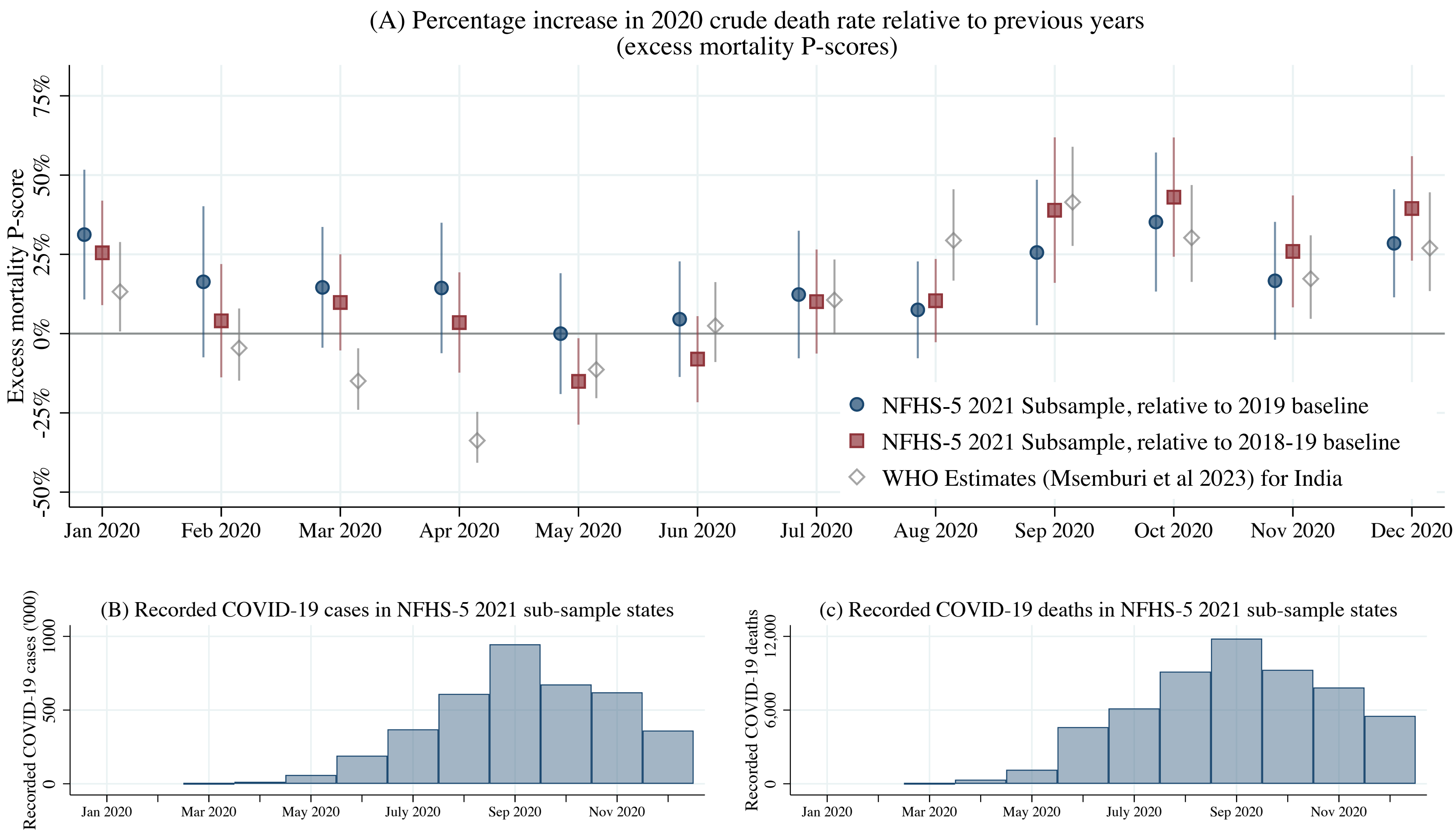
Note: The figure shows the decomposition of the change in life expectancy between 2019 and 2020 into contributions from changes in mortality in different age groups based on Arriaga's decomposition for (A) females and (B) males (Arriaga, 1984). Estimates are for the NFHS-5 2021 subsample and use sample weights. 95% confidence intervals are shown in parentheses below estimates and are calculated using a cluster-bootstrap approach. Source: NFHS-5.

Figure 4: Greater declines in period life expectancy between 2019 and 2020 for disadvantaged social groups compared to privileged social groups



Note: The figure shows life expectancy at birth in 2019 and 2020 for the (A) combined female and male population, (B) females, and (C) males, separately. Life expectancy is calculated based on standard life table procedures. Estimates are for the NFHS-5 2021 subsample and use sample weights. The vertical lines around each estimate represent 95% confidence intervals calculated using a cluster-bootstrap approach. See Figure SI-7 for estimates of the decline in life expectancy between the two years. Source: NFHS-5.

Figure 5: Greater percentage increase in mortality in the last four months of 2020 relative to prior years



Note: (A) shows excess mortality estimated in the NFHS-5 2021 subsample and compares it to excess mortality estimated by the WHO (Msemburi et al., 2023). For the NFHS-5 2021 subsample, monthly excess mortality P-scores are estimated based on the procedure described in the Methods section. The percentage increase in monthly crude death rates in 2020 is shown relative to two different baselines: 2019 only, and the average of mortality in 2018 and 2019. NFHS-5 estimates use sample weights, and 95% confidence intervals calculated using a cluster-bootstrap approach are shown as vertical around estimates. WHO estimates are extrapolated for India based on data from 17 states and union territories. 95% confidence intervals are shown as vertical lines, and were provided in Msemburi et al. (2023). (B) COVID-19 cases and (C) COVID-19 deaths, reported by labs and hospitals to state agencies. Sources: NFHS-5, Msemburi et al. (2023), and covid19india.org (2023).

Supplementary Information

Description of SI tables and Figures

Table SI-1 shows that the subsample of households interviewed by the NFHS-5 in 2021 is similar in characteristics to the full NFHS-5 sample. Table SI-2 shows that alive and dead individuals observed in the subsample are similar in terms of age and sex as the full sample.

Table SI-3 displays SARS-CoV-2 seroprevalence for India's states as measured by India's third national serosurvey, conducted between December 2020 and January 2021 (Murhekar et al., 2021). Seroprevalence at the state level is calculated as the mean prevalence in the districts of respective state. This serosurvey, which was conducted at the end of the period for which we report excess mortality estimates from the NHFS-5 subsample, does not indicate that seroprevalence in the subsample states was different from the states not in the subsample.

Table SI-4 calculates implied excess deaths for all of India, based on excess mortality observed during the pandemic months in the NFHS-5 subsample. We estimate 1.18 million excess deaths in 2020. The table also documents a female disadvantage in excess deaths.

Figure SI-1 shows the distribution of NFHS-5 interview by survey phase. Most Phase 1 interviews were conducted between July 2019 and December 2019. Phase 2 interviews began before the first national lockdown and were conducted largely between January and March 2020. Interviews resumed in October 2020, but the vast majority of interviews were conducted between December 2020 and April 2021. Interviews in January 2021 and after, shown as blue bars, comprise our analytical subsample.

Figure SI-2 shows the distribution of NFHS-5 interviews by state. Related to this, Figure SI-3 shows the spatial distribution of Primary Sampling Units (PSU) in our subsample in blue. PSUs that were interviewed before 2021 are shown in red.

Figure SI-4 compares excess mortality estimates from our subsample to excess mortality estimates from hypothetical sub-subsamples which exclude one subsample state each. We created 14 sub-subsamples. Excess mortality estimates observed in these hypothetical sub-subsamples is not statistically different from excess mortality in the subsample.

Figure SI-5 shows annualized age-standardized crude death rates in the sub-sample by calendar month for the years 2018, 2019, and 2020. Pooled, rural, and urban mortality estimates are shown separately. Given that a larger share of the subsample consists of rural residents, confidence intervals around urban estimates are larger than those around rural estimates. In rural areas and in the overall subsample, we observe elevated mortality towards the end of the year. In urban areas, we observe elevated mortality in October 2020. Figure 5 shows that confirmed COVID-19 cases and deaths peaked in September 2020. A surge in cases around September 2020 is consistent with a mortality peak slightly later in October 2020. Thus, one interpretation from Figure SI-5 is that case and death data tracked the epidemic in urban areas better. Data on confirmed COVID-19 cases and deaths fails to capture elevated mortality in rural areas towards the end of the year.

In Figure SI-6, the y-axis shows age-standardized and annualized crude death rates, and the x-axis shows the number of months prior to the interview month in NFHS-5 Phase 1. NFHS-5 Phase 1 interviews were conducted in the second half of 2019 and in early 2020. We do not detect systematic differences in reported crude death rates between calendar months more proximate in the past to interview months and calendar months further in the past to interview months.

Figure SI-7 examines declines in life expectancy between 2019 and 2020. Estimates are shown by sex, social group, and the pooled subsample. Declines are highest among Muslims. They are also higher among females relative to males; and among other marginalized groups, such as Scheduled Castes and Scheduled Tribes, relative to high caste Hindus.

Figure SI-8 shows that although overall declines in rural and urban areas are similar, declines in life expectancy between 2019 and 2020 are larger among rural males, rural females, and urban females compared to urban males.

Figure SI-9 shows results from an Arriaga decomposition, decomposing the age-specific contribution (for abridged life table age groups) of changes in life expectancy between 2019 and 2020 for male and females separately. For females, we see a large contribution of childhood ages (0-4), young adulthood (15-24), and older ages (50+). For males, we see a large contribution of childhood ages, and also of older working ages (40-59).

Figure SI-10 examines if life expectancy declines by social group differ for rural and urban residents. Muslims in urban areas experienced the largest declines in life expectancy. On the other hand, urban high caste Hindus, the most privileged group in Indian society, experienced the smallest declines in life expectancy between 2019 and 2020.

Table SI-1: Comparison of NFHS-5 2021 subsample (the main analysis sample) with the full NFHS-5 sample on household socio-economic characteristics

	Subsample (Only 2021 interviews)	Full sample	Difference
	(1)	(2)	(1) - (2)
Rural	0.67	0.67	-0.00
	(0.01)	(0.00)	(0.01)
Religion			
Hindu	0.84	0.81	0.03
	(0.00)	(0.00)	(0.00)
Muslim	0.10	0.13	-0.03
	(0.00)	(0.00)	(0.00)
Christian	0.02	0.03	-0.01
	(0.00)	(0.00)	(0.00)
Sikh	0.03	0.01	0.01
	(0.00)	(0.00)	(0.00)
Other	0.01	0.01	-0.01
	(0.00)	(0.00)	(0.00)
Caste			
Scheduled Caste	0.23	0.23	0.00
	(0.00)	(0.00)	(0.00)
Scheduled Tribe	0.11	0.10	0.01
	(0.00)	(0.00)	(0.00)
Other Backward Class	0.48	0.43	0.06
	(0.00)	(0.00)	(0.01)
Other	0.18	0.25	-0.07
	(0.00)	(0.00)	(0.00)
Wealth Index			
Poorest	0.22	0.20	0.02
	(0.00)	(0.00)	(0.00)
Poorer	0.19	0.20	-0.02
	(0.00)	(0.00)	(0.00)
Middle	0.18	0.21	-0.03
	(0.00)	(0.00)	(0.00)
Richer	0.18	0.20	-0.02
	(0.00)	(0.00)	(0.00)
Richest	0.23	0.19	0.04
	(0.00)	(0.00)	(0.01)

Notes: Column (1) of the table gives summary statistics of the NFHS-5 2021 subsample. Column (2) displays summary statistics using the full NFHS-5 sample. Column (3) reports the difference between the two. The wealth index is computed by The DHS Program, and is based on household ownership of assets such as motorcycle or television; dwelling characteristics such as flooring material; type of drinking water source; toilet facilities; and type of cooking fuel. Estimates use sample weights. Clustered standard errors are reported in parentheses under each estimate. Source: NFHS-5.

Table SI-2: Comparison of NFHS-5 2021 subsample (the main analysis sample) with the full NFHS-5 sample on individual demographic characteristics

	Subsample (Only 2021 interviews)	Full sample	Difference
	(1)	(2)	(1) - (2)
Female	0.50	0.50	-0.00
	(0.00)	(0.00)	(0.00)
Age Groups			
0-5	0.15	0.13	0.02
	(0.00)	(0.00)	(0.00)
6-20	0.26	0.26	0.00
	(0.00)	(0.00)	(0.00)
21-40	0.29	0.30	-0.01
	(0.00)	(0.00)	(0.00)
41-60	0.20	0.21	-0.01
	(0.00)	(0.00)	(0.00)
61-80	0.09	0.09	-0.00
	(0.00)	(0.00)	(0.00)
80+	0.01	0.01	0.00
	(0.00)	(0.00)	(0.00)
Population Coverage(%)	23.16	100	–
N	766,209	3,002,140	–

Notes: Column (1) of the table gives summary statistics of the NFHS-5 2021 subsample. Column (2) displays summary statistics using the full NFHS-5 sample. Column (3) reports the difference between the two. Age groups are constructed using the last observed age of individuals. For deceased individuals, this is the age at death, whereas for those who were alive, it is the age at the time of the interview. Estimates use sample weights. Clustered standard errors are reported in parentheses under each estimate. Source: NFHS-5.

Table SI-3: Seroprevalence from India's third national serosurvey at the end of 2020 is similar in NFHS-5 2021 subsample states and states interviewed by NFHS before 2021

	Seroprevalence (%) (Dec 2020 - Jan 2021)	Population (000) (2020 estimate)
States in NFHS-5 2021 Subsample		
Chhattisgarh	24.2	29,109
Haryana	21.0	29,077
Jharkhand	21.3	37,937
Madhya Pradesh	22.8	83,374
Odisha	33.9	45,350
Punjab	19.8	30,099
Rajasthan	27.5	78,273
Tamil Nadu	28.8	76,049
Uttar Pradesh	23.3	227,943
Uttarakhand	12.0	11,270
Estimated seroprevalence in subsample	24.6	
States interviewed by NFHS-5 before 2021		
Andhra Pradesh	35.1	52,504
Assam	32.1	34,668
Bihar	29.5	121,302
Gujarat	22.1	68,862
Himachal Pradesh	46.5	7,347
Jammu and Kashmir	32.0	13,305
Karnataka	29.3	66,322
Kerala	9.7	35,307
Maharashtra	21.6	123,295
Telangana	29.5	37,473
West Bengal	28.1	97,516
Estimated seroprevalence out of subsample	26.8	

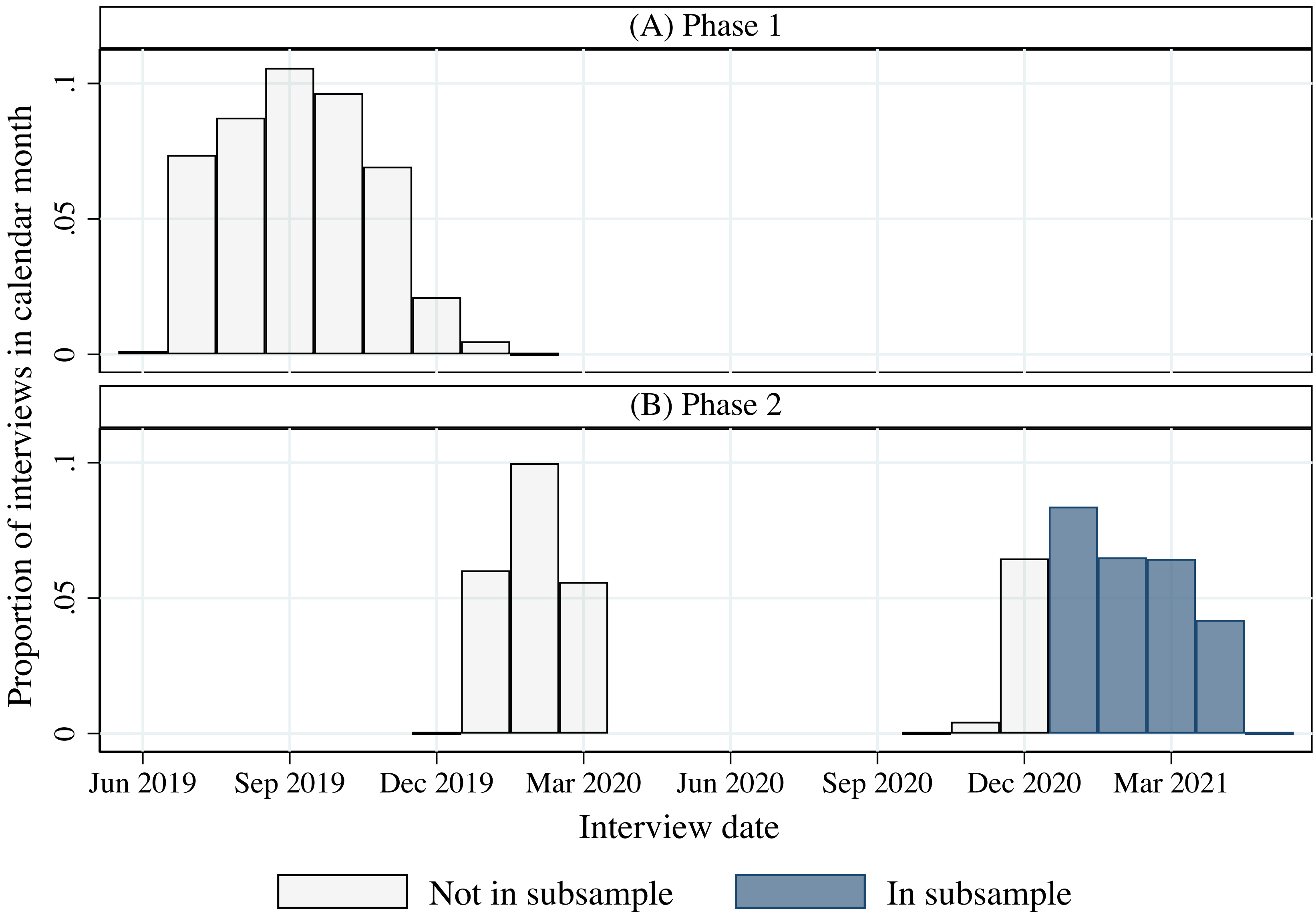
Note: Seroprevalence is based on estimates of India's third national serosurvey, conducted in December 2020 and January 2021 (Murhekar et al., 2021). Estimates for Union Territories are not available from the serosurvey. Overall seroprevalence was estimated to be 24.1% [95% CI: 23.0-25.3]. Population estimates for 2020 are from population projections by India's National Commission of Population (National Commission on Population, 2019). Overall seroprevalence in NFHS-5 2021 subsample and out-of-subsample states is estimated as a weighted average of state-level seroprevalence and population estimates. Sources: Murhekar et al. (2021), National Commission on Population (2019).

Table SI-4: National estimates of excess mortality in India

	Female (1)	Male (2)	Pooled (3)
Observed excess mortality P-scores, (%)	21.2	14.1	17.0
(Apr - Dec 2020, compared to 2019 baseline)	[11.0, 31.4]	[5.2, 22.9]	[10.6, 23.3]
Annual number of deaths, 2019 (millions)	4.28	5.00	9.28
(UN WPP Estimate)			
Implied excess deaths nationally (millions)	0.68	0.53	1.18
(Apr - Dec 2020, compared to 2019 baseline)	[0.35, 1.01]	[0.20, 0.86]	[0.74, 1.63]

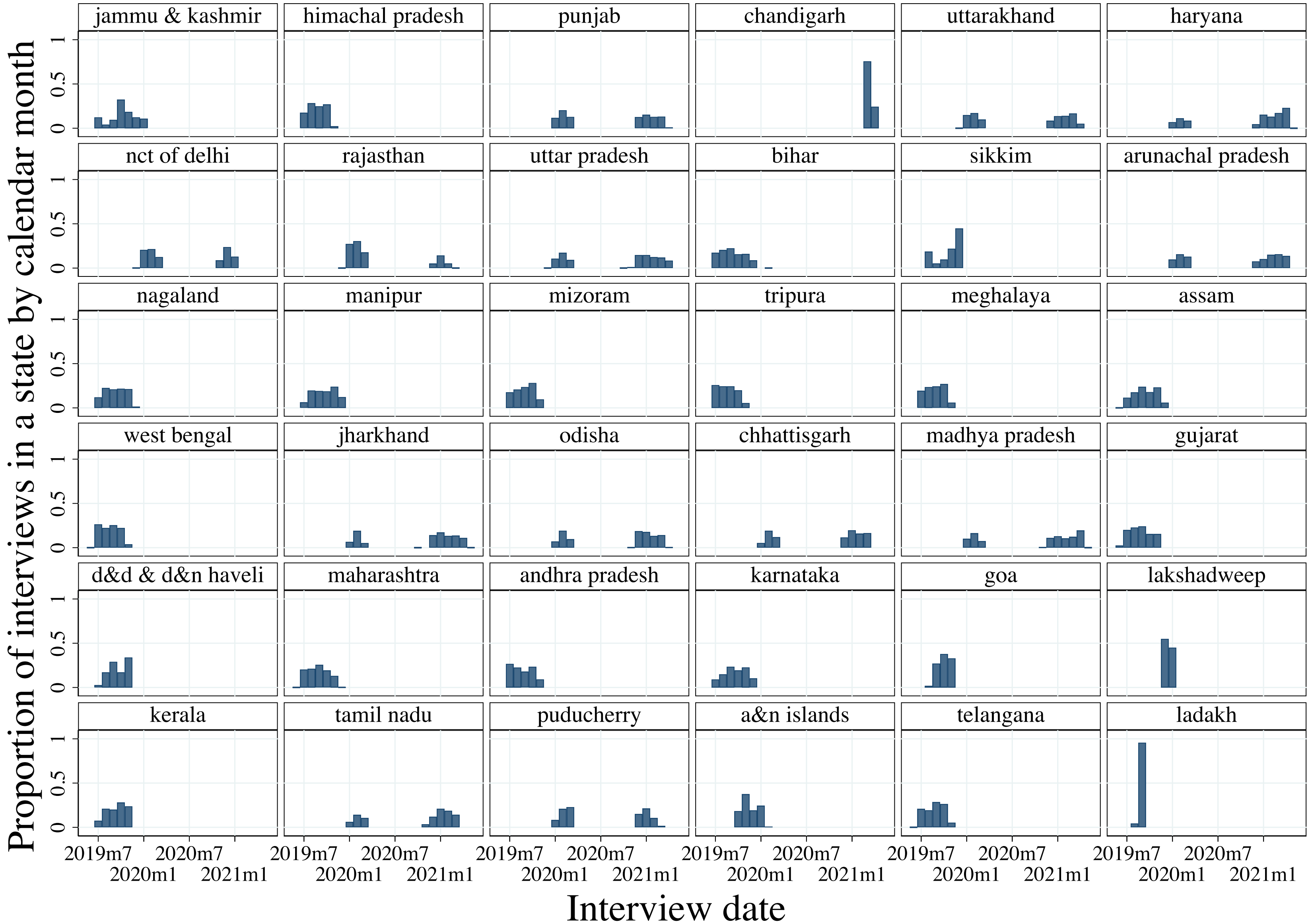
Note: Row 1 of the table shows observed percentage increase in mortality in April through December 2020 (the 2020 pandemic months), relative to the same months in 2019, in the NFHS-5 2021 subsample. Row 2 shows annual estimates of deaths in India nationally in 2019 from the UN World Population Prospects (United Nations, 2022). Row 3 shows the implied excess deaths nationally if excess mortality from the NFHS-5 2021 subsample were observed nationally. Row 3 is calculated based on the following formula: $\frac{row1 \times row2 \times 0.75}{100}$. 95% confidence intervals for NFHS-5 estimates are shown in brackets below estimates and are calculated using a cluster-bootstrap approach. Sources: NFHS-5, UN WPP 2022.

Figure SI-1: Proportion of NFHS-5 interviews conducted in each calendar month in each survey phase



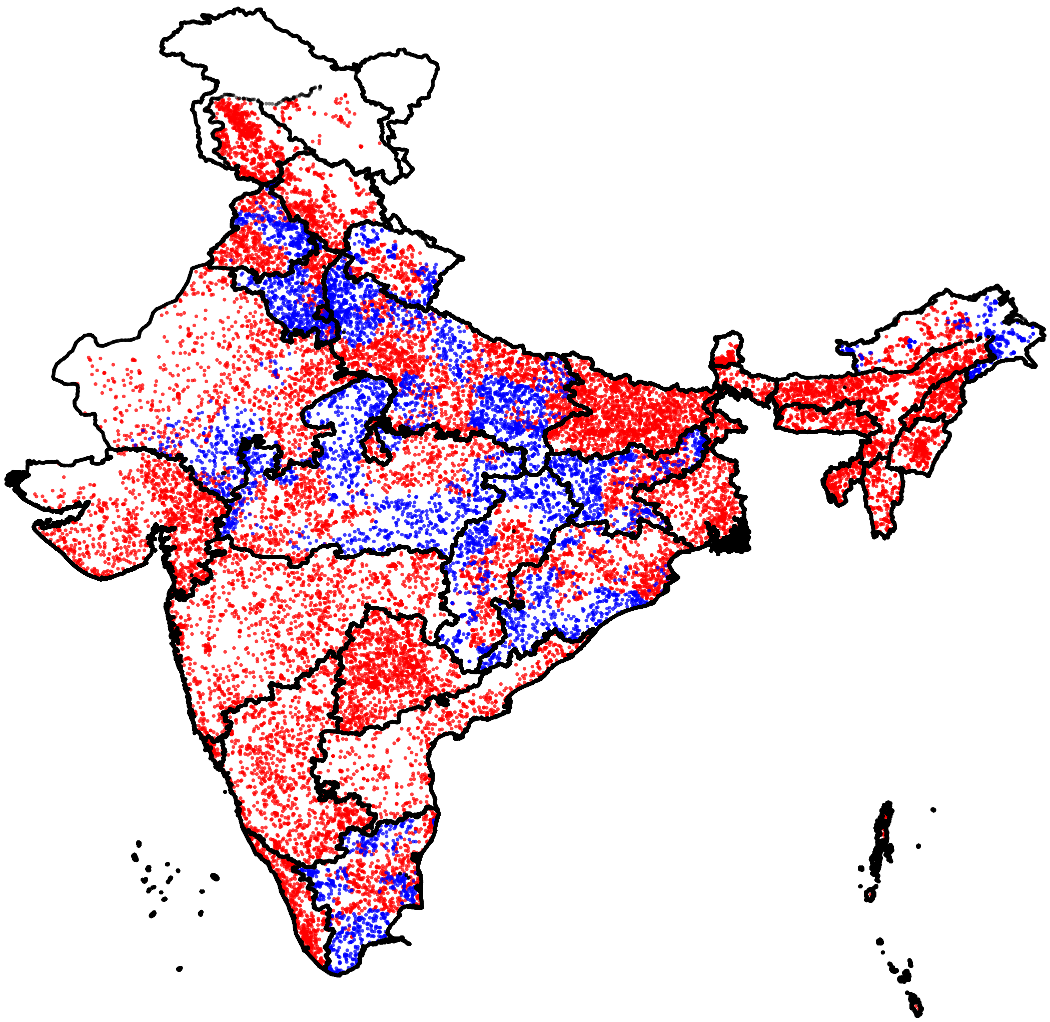
Note: The figure shows the proportion of NFHS-5 interviews carried out in each calendar month in (A) Phase 1 and (B) Phase 2. The main analysis sample used in this paper includes households interviewed in 2021, shown in blue. Source: NFHS-5.

Figure SI-2: Proportion of NFHS-5 interviews conducted in each calendar month in each state or union territory between 2019 and 2021



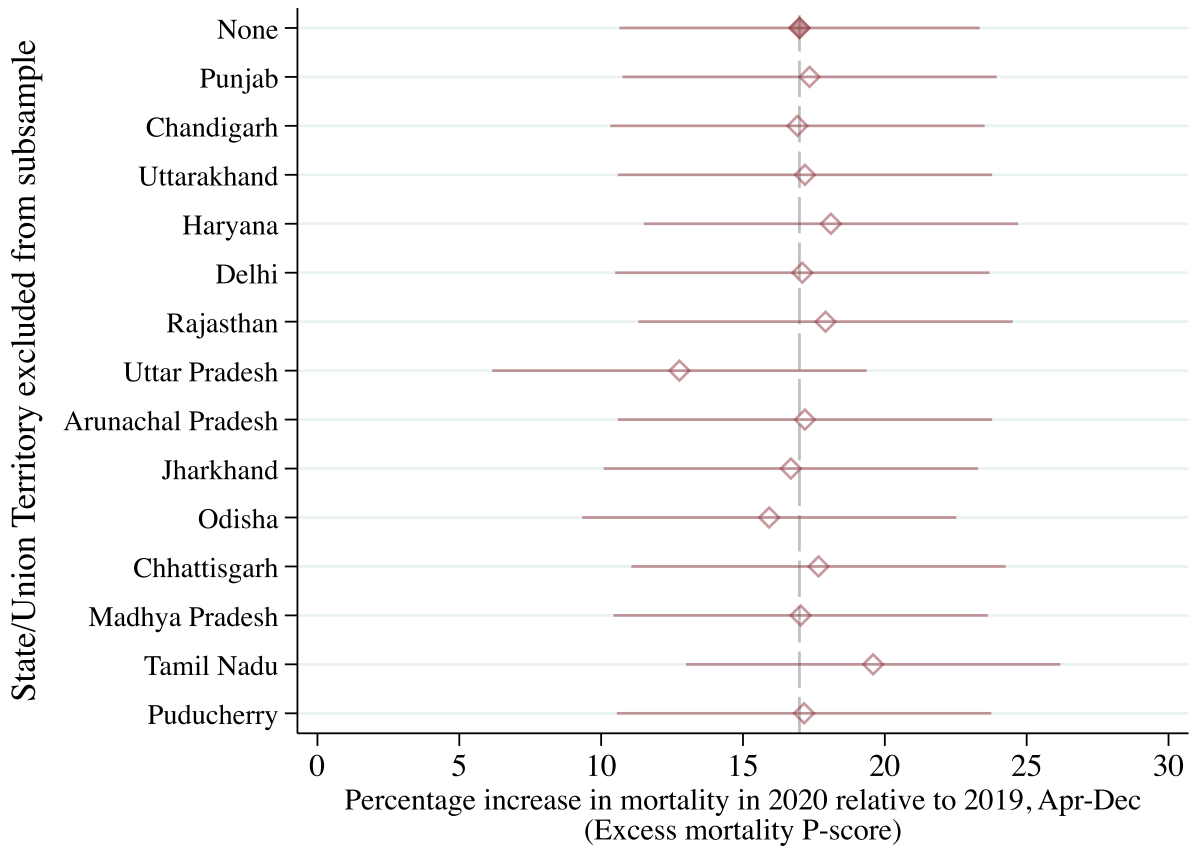
Note: The figure shows the proportion of NFHS-5 interviews carried out in each calendar month in each state. The main analysis sample used in this paper includes households interviewed in 2021. Source: NFHS-5.

Figure SI-3: Spatial distribution of Primary Sampling Units (Clusters) in NFHS-5 2021 subsample



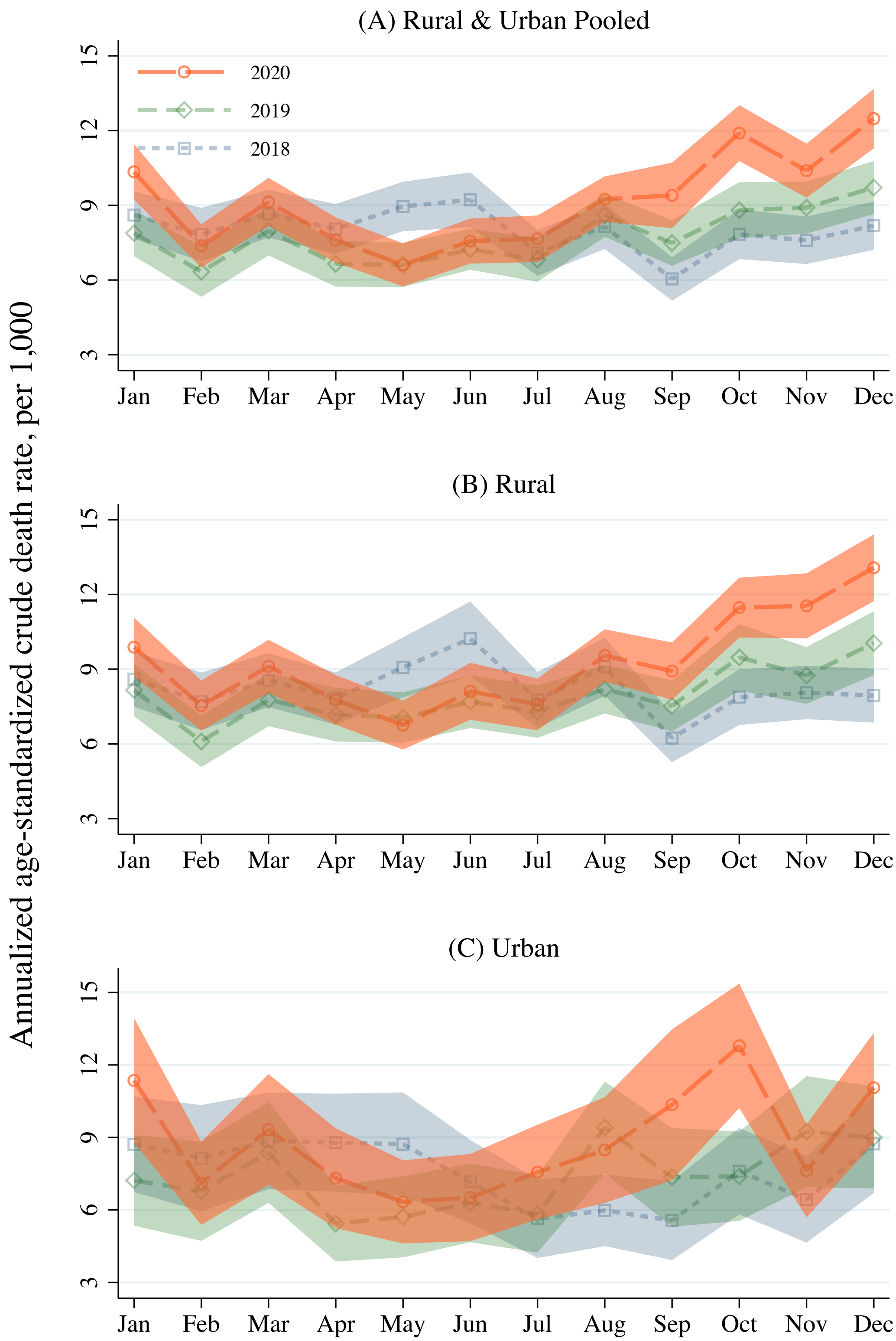
Note: Primary Sampling Units in the NFHS-5 2021 subsample are shown in blue, and Primary Sampling Units interviewed in 2019 and 2020 (and therefore not in the NFHS-5 2021 subsample), are shown in red. Source: NFHS-5.

Figure SI-4: Robustness check: Excess mortality estimate is robust to excluding one state at a time in hypothetical sub-subsamples



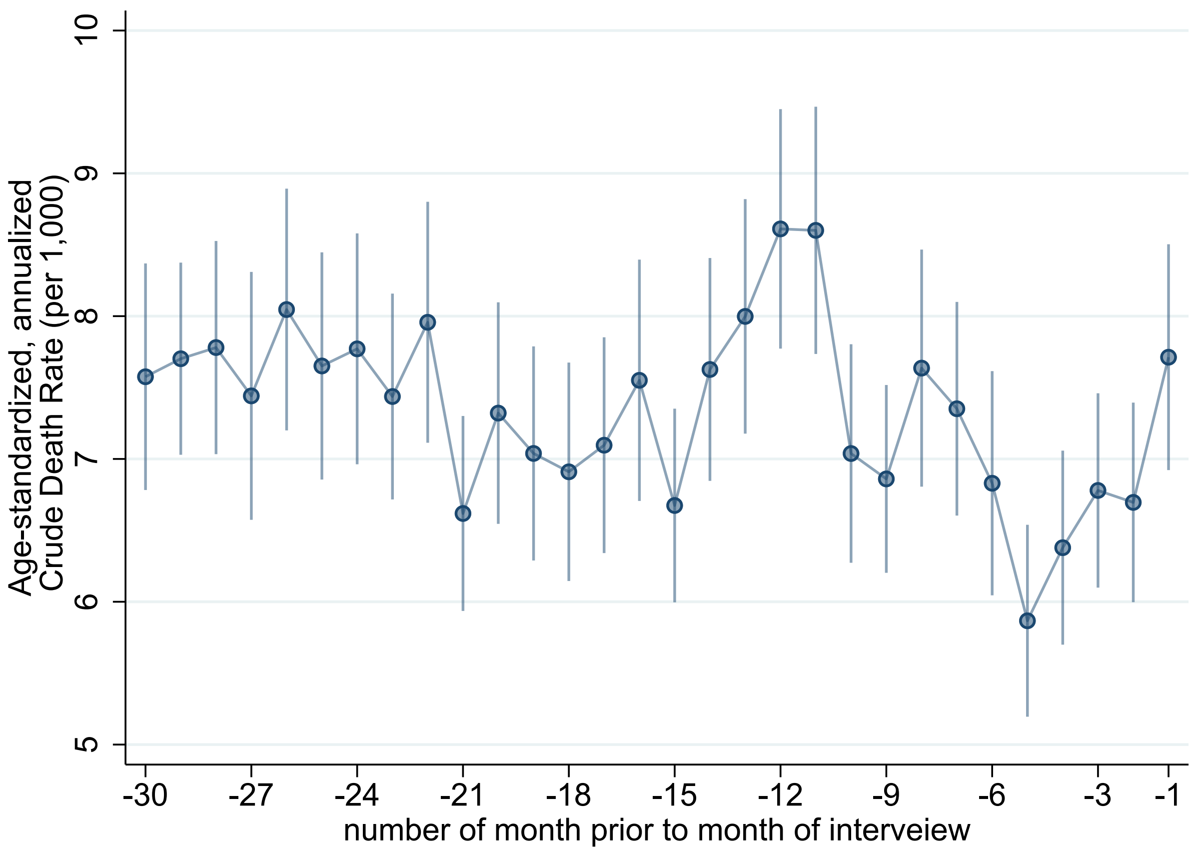
Note: The figure shows results from an analysis which estimates the excess mortality P-score from hypothetical sub-subsamples (subsamples of the NFHS-5 2021 subsample) that exclude one state at a time from the subsample. Excess mortality P-scores are the percentage increase in mortality in April through December 2020 relative to the same months in 2019. Estimates use sample weights, and 95% confidence intervals calculated using a cluster-bootstrap approach are shown as horizontal lines around estimates. Source: NFHS-5.

Figure SI-5: Crude death rates by calendar month, NFHS-5 2021 interviews subsample



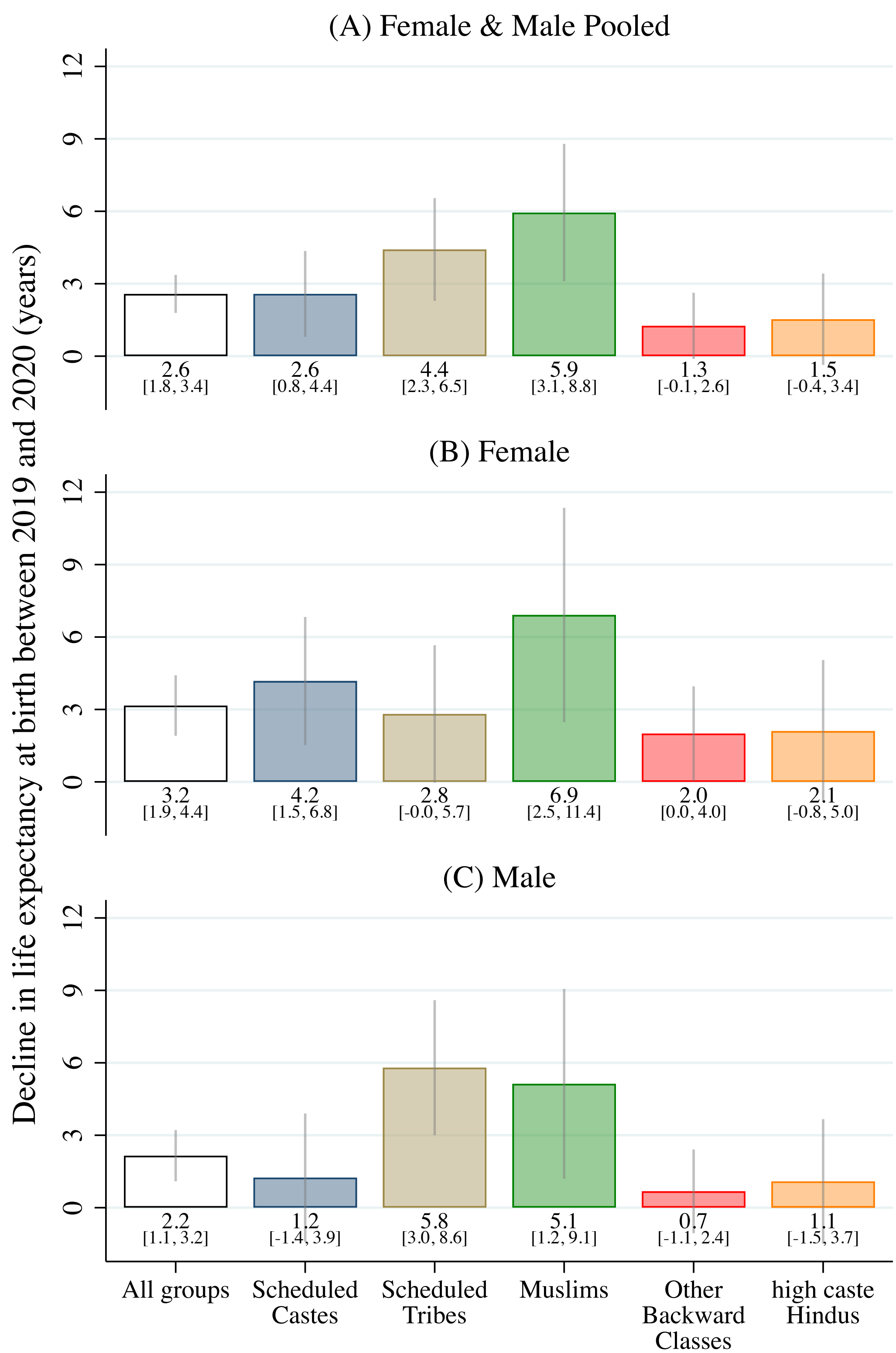
Note: Monthly crude death rates for (A) the combined rural and urban population, (B) the rural population, and (C) the urban population, separately. Estimates are for the NFHS-5 2021 subsample and use sample weights. Monthly crude death rates are age-standardized and annualized. 95% confidence intervals are shown as the shaded area around estimates and are calculated using a cluster-bootstrap approach. Source: NFHS-5.

Figure SI-6: Death reporting is not biased towards months more proximate to the interview



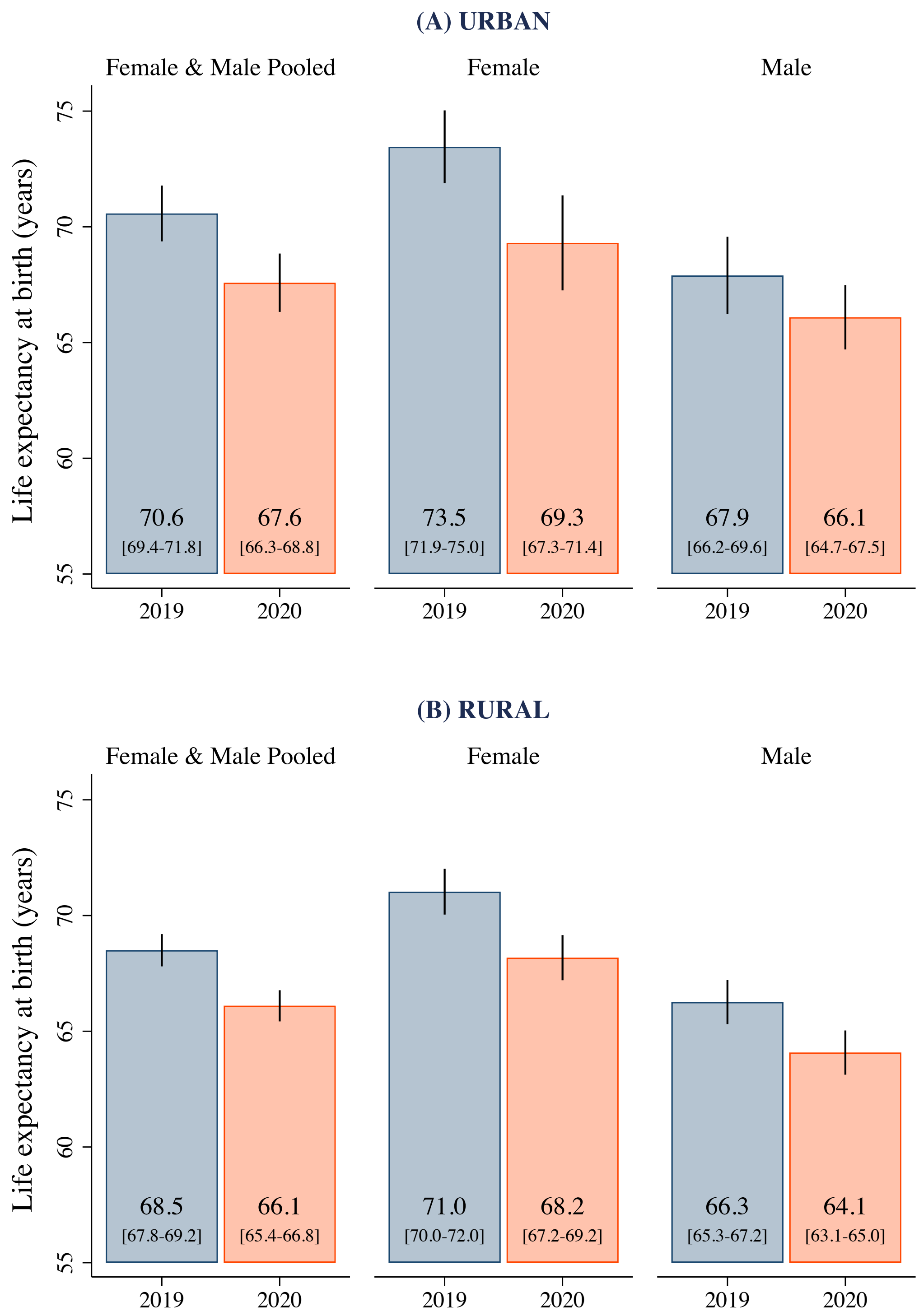
Note: The figure displays monthly crude death rates, in months relative to the interview date. Estimates are for the NFHS-5 2019 subsample, which interviewed households between June and December 2019. Estimates use sample weights. Monthly crude death rates are age-standardized and annualized. The vertical lines around each estimate represent 95% confidence intervals calculated using a cluster-bootstrap approach. Source: NFHS-5.

Figure SI-7: Decline in period life expectancy at birth between 2019 and 2020 by sex and social group



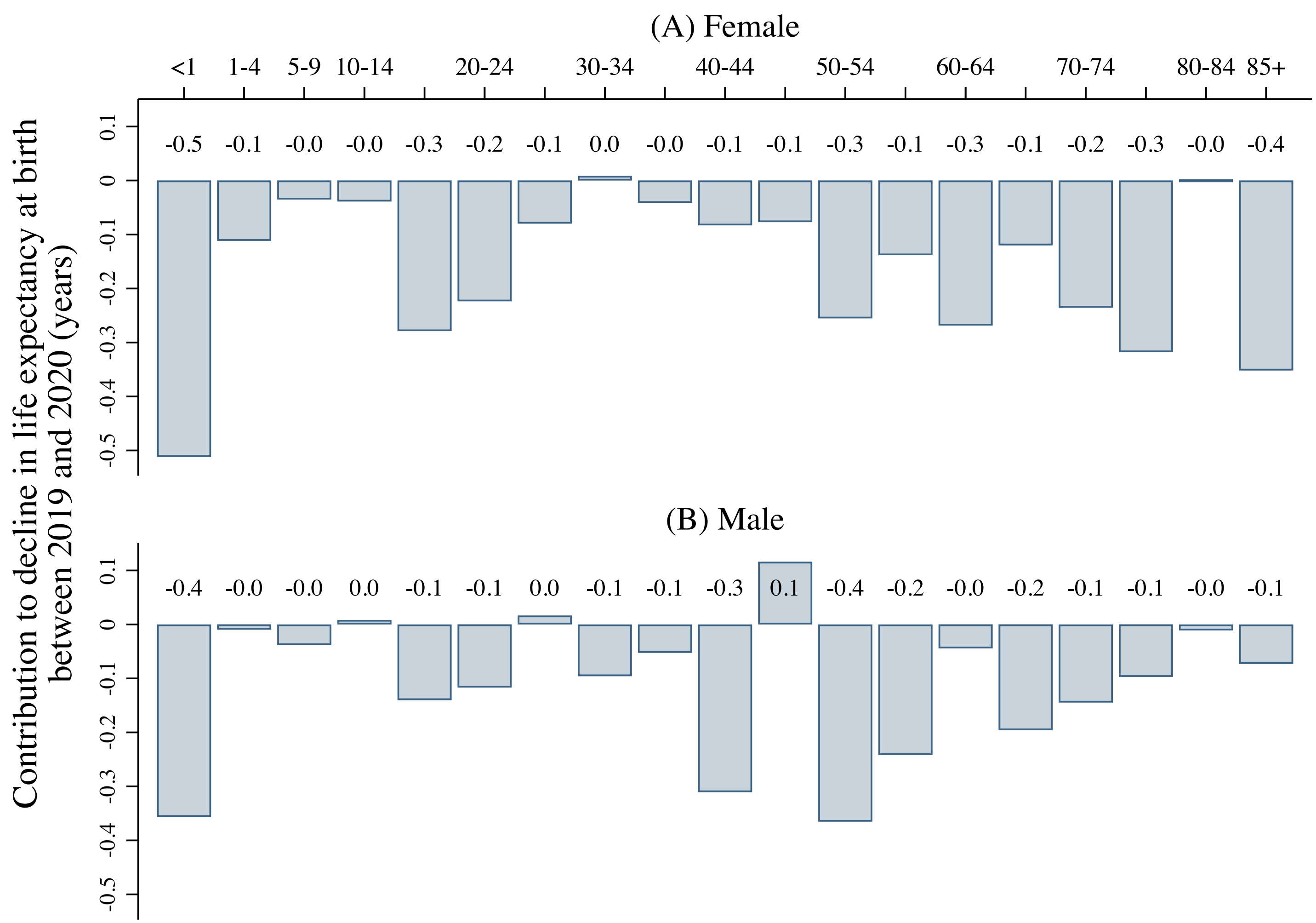
Note: The figure shows declines in life expectancy at birth by social group between 2019 and 2020 for (A) the combined female and male population, (B) females, and (C) males, separately. Life expectancy is calculated based on standard life table procedures. Estimates are for the NFHS-5 2021 subsample and use sample weights. The vertical lines around each estimate represent 95% confidence intervals calculated using a cluster-bootstrap approach. Source: NFHS-5.

Figure SI-8: Period life expectancy at birth in 2019 and 2020 by sex and residence



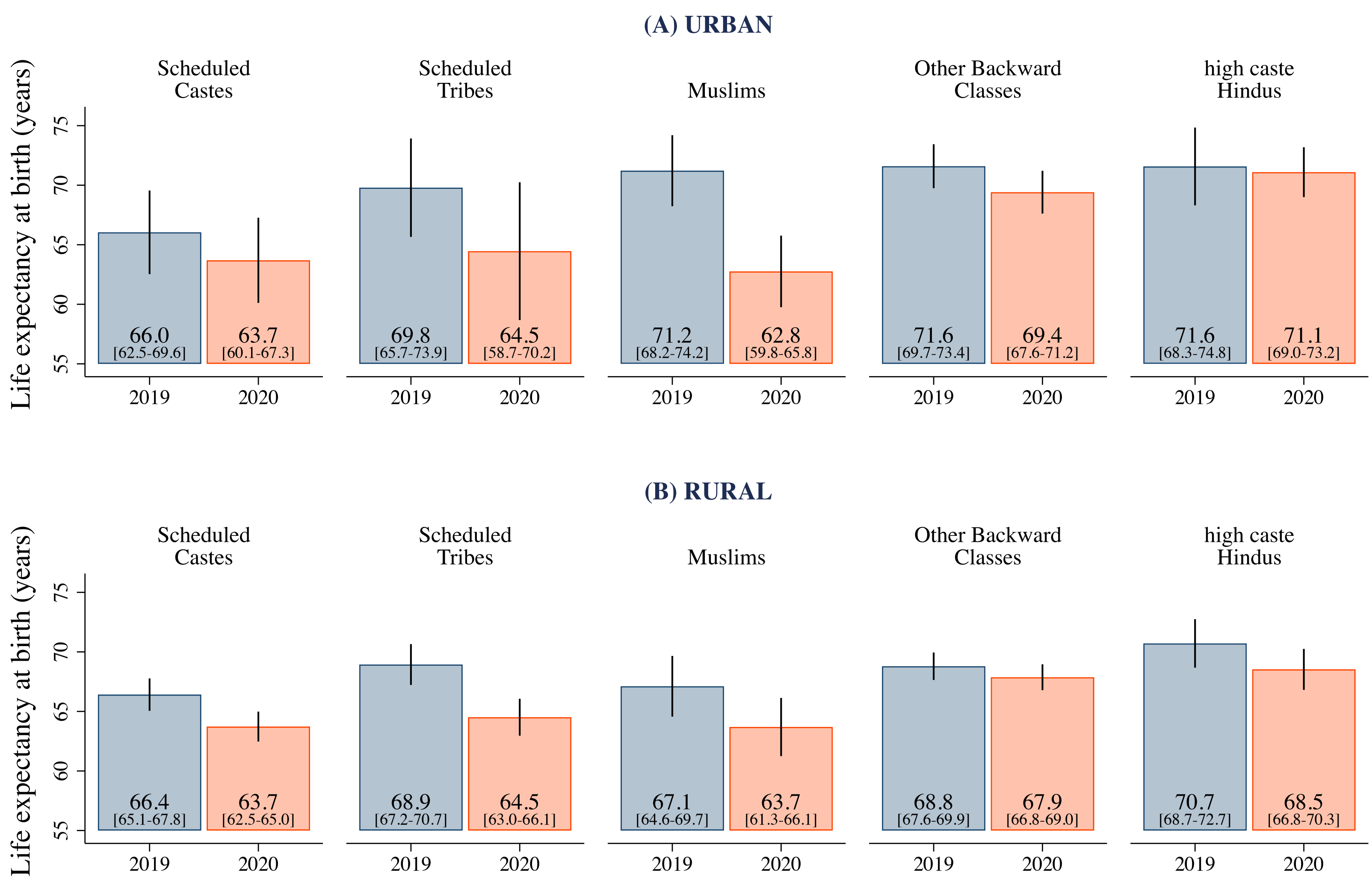
Note: The figure shows life expectancy at birth in 2019 and 2020 for (A) the urban population and (B) the rural population. In each panel, estimates are shown for the combined female and male population, for females, and males, separately. Life expectancy is calculated based on standard life table procedures. Estimates are for the NFHS-5 2021 subsample and use sample weights. The vertical lines around each estimate represent 95% confidence intervals calculated using a cluster-bootstrap approach. Source: NFHS-5.

Figure SI-9: Age-specific contributions (by abridged life table age groups) to changes in period life expectancy at birth between 2019 and 2020



Note: The figure shows the decomposition of the change in life expectancy between 2019 and 2020 into contributions from changes in mortality in different age groups based on Arriaga's decomposition for (A) females and (B) males (Arriaga, 1984). Estimates are for the NFHS-5 2021 subsample and use sample weights. Source: NFHS-5.

Figure SI-10: Period life expectancy at birth in 2019 and 2020 by social group and residence



Note: The figure shows life expectancy at birth in 2019 and 2020 for (A) the urban population and (B) the rural population. In each panel, estimates are shown for each social group separately. Life expectancy is calculated based on standard life table procedures. Estimates are for the NFHS-5 2021 subsample and use sample weights. The vertical lines around each estimate represent 95% confidence intervals calculated using a cluster-bootstrap approach. Source: NFHS-5.