

Mishra, Shruti; Goli, Srinivas; Rammohan, Anu

Working Paper

Mortality Selection Bias and the Underestimation of Health Inequalities in Later Life: Evidence from India

Suggested Citation: Mishra, Shruti; Goli, Srinivas; Rammohan, Anu (2026) : Mortality Selection Bias and the Underestimation of Health Inequalities in Later Life: Evidence from India, ZBW - Leibniz Information Centre for Economics, Kiel, Hamburg

This Version is available at:

<https://hdl.handle.net/10419/340190>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Mortality Selection Bias and the Underestimation of Health Inequalities in Later Life: Evidence from India

Shruti Mishra¹, Srinivas Goli² and Anu Rammohan³

ABSTRACT

Health inequalities between socioeconomic groups tend to narrow at older ages, in studies using conventional cross-sectional data. This apparent convergence, however, is largely an artifact of mortality selection bias: the systematic dropout of the most disadvantaged individuals from the population before they reach old age. Our study quantifies the extent of this bias in India and provides corrected estimates of age-specific health inequality using pooled data from three rounds of the National Sample Survey, comprising 1,278,372 individuals and 6,630 deaths. We document the wealth-mortality gradient using survival analysis and concentration indices, and apply Inverse Probability Weighting (IPW) to adjust health outcomes for differential survival, creating a counterfactual population that accounts for selection on observed characteristics. Finally, we implement Deaton-Paxson age-period-cohort (APC) models with and without selection adjustment. Our analysis finds evidence of a wealth gradient in mortality among working-age adults (Hazard Ratio (HR) for rich vs. poor: 0.576, $p < 0.01$ for ages < 65), which disappears after age 65 (HR: 0.885, $p > 0.10$). Unadjusted age-health profiles flatten at older ages. However, after IPW correction, every age coefficient increases relative to unadjusted estimates, indicating systematic underestimation of poor health at all ages. Mortality selection bias significantly underestimates the true extent of poor health among lower SES groups in India, masking up to 83% of the health burden among young adults. Correcting this bias provides a more accurate picture of population health with direct implications for resource allocation and targeting interventions toward vulnerable populations.

Keywords: Mortality Selection, Health Inequality, Inverse Probability Weighting, Age-Period-Cohort Analysis, India, Economics and Human Biology

JEL Classification: I14, I15, J11, C23

¹Post-Doctoral Fellow, International Institute for Population Sciences, Mumbai; Assistant Professor, Department of Economics, Ramdayalu Singh College, Babasaheb Bhimrao Ambedkar Bihar University, Muzaffarpur, Bihar, India; E-mail: shrutibhu28@gmail.com

²Associate Professor, Department of Fertility & Social Demography, International Institute for Population Sciences, Mumbai, Maharashtra, India; E-mail: srinivasgoli@iipsindia.ac.in

³ Professor of Economics and Director of International Relations at the Business School, The University of Western Australia (M251), 35 Stirling Highway, Perth, Australia; E-mail: anu.rammohan@uwa.edu

1. Introduction

Socioeconomic disparities in health and survival are among the most persistent features of population heterogeneity. Individuals with fewer resources experience higher rates of morbidity and premature mortality than their more affluent peers (Marmot et al. 1984; Adler and Rehkopf 2008; Chetty et al. 2016). Yet analyses based on cross-sectional data often reveal a puzzling pattern: the health gradient by socioeconomic status (SES) appears to weaken, or even reverse, with age (Beckett, 2000; House et al., 1994; Ferraro & Farmer, 1996; Lynch, 2003; Dupre, 2007; Hoffmann, 2011). The poor, though worse off earlier in life, seem to “catch up” in later years (Elo & Preston, 1996; Beckett, 2000; Herd, 2006). This apparent convergence has often been interpreted as evidence that the physiological processes of aging eventually dominate social advantage, a view encapsulated in the *age-as-leveller hypothesis* (Beckett 2000; Lynch, 2003; Dannefer & Sell, 1988; O’Rand, 2003; Ross & Wu, 1996).

A central tenet of demographic analysis, however, is that the composition of the living population changes continuously over time through births, deaths, and migration. When mortality differs systematically across social strata, survival itself becomes a selective process. The individuals who survive to old age are not random members of their birth cohorts but are disproportionately drawn from the healthiest and most resilient subgroups. This mortality selection alters the composition of cohorts as they age, such that the frailest individuals in disadvantaged groups die earlier, leaving behind those with better underlying health. In contrast, advantaged groups retain both robust and frail members because of lower mortality. Consequently, cross-sectional comparisons of health by SES at older ages conflate *selection* with *change*, producing the statistical illusion of health convergence that does not reflect actual equality in health endowments or exposures (Vaupel, Manton, and Stallard 1979; Palloni and Ewbank 2004).

This compositional shift has strong implications for policy and measurement. If policy analysts interpret this apparent narrowing of disparities as a natural feature of the aging process, they may underestimate the cumulative burden of disadvantage that persists throughout the life course. Similarly, projections of future health and disability among older populations may be biased downward, especially for disadvantaged groups selectively depleted by premature mortality. From a demographic perspective, ignoring selection implies that the observed elderly population is taken to represent the health experience of its birth cohort, when in fact it reflects only the surviving, and therefore selected, subset of that cohort.

Despite its importance, empirical evidence quantifying mortality selection bias in health inequality is limited. In high-income settings, longitudinal studies have linked attrition and survival selection to flattened health gradients at older ages (Beckett 2000; Weuve et al. 2012). Evidence from low and middle-income countries (LMICs) is scarcer, though the mechanisms of selection are likely more pronounced where mortality differentials are wide and vital registration is incomplete. In India, for example, a rapid demographic transition has unfolded alongside persistent disparities in income, education, caste, and residence. National surveys document

strong socioeconomic gradients in morbidity and longevity, but most remain cross-sectional and therefore implicitly assume representative survival. As a result, the degree to which selective mortality shapes the observed age pattern of health inequality remains unknown.

This study combines demographic and econometric methods to quantify how mortality selection affects estimates of health inequality across the lifespan. Using pooled microdata from three rounds of the *National Sample Survey* (NSS), our analysis provides the first national assessment of mortality selection bias in health for India.

This paper contributes to demographic research on health and inequality in four ways. Firstly, it quantifies the demographic selection mechanism by providing the first national-level estimate of mortality selection bias in health inequality for India, demonstrating how differential survival shapes observed patterns of morbidity across the life course. Secondly, it applies inverse probability weighting (IPW) to reconstruct the health distribution of the population prior to selective mortality, thereby linking mortality gradients to age-specific health outcomes. Thirdly, we estimate age-period-cohort (APC) models with and without selection correction to assess how mortality selection alters age-health patterns. The application of IPW within an APC design shows how selection bias can be corrected even without longitudinal follow-up data, offering a replicable framework for other LMIC contexts. Finally, the findings from this analysis have implications for measurement and policy. The findings show that conventional cross-sectional data significantly underestimate the true burden of poor health among lower-SES adults. Accounting for mortality selection alters conclusions about aging, inequality, and the demographic shape of health, refining projections critical to health-system planning and social protection policy. Furthermore, by situating this analysis in a rapidly transitioning developing country, we demonstrate that mortality selection biases are even more consequential in populations undergoing rapid epidemiological and economic change, extending a literature that has predominantly focused on developed settings.

In the Indian context, mortality selection is likely to be especially influential as mortality decline has occurred unevenly across different regions and socioeconomic groups. Individuals from lower-SES experience higher mortality at nearly all ages. However, as data generally captures only the survivors, health estimates drawn from cross-sectional surveys may therefore systematically underestimate the prevalence of ill-health among those socioeconomically disadvantaged subgroups whose members are most likely to die before the time of observation. Given that India's older population is projected to more than double between 2020 and 2050, failing to adjust for selective survival could lead to biased assessments of aging, dependency, and future healthcare demands.

To address these gaps, we correct for mortality selection to cross-sectional microdata from India, integrating demographic and econometric approaches. We then embed this reweighted data within a Deaton-Paxson age-period-cohort (APC) framework (Deaton and Paxson 1994) to separate life-cycle, cohort, and period components of health inequality, with and without correction for selection.

Ultimately, this paper argues that what appears to be an age-related convergence in health inequality in India is, in large part, an artifact of mortality selection bias.

1.1 Conceptual Framework

The theoretical foundation for mortality selection lies in models of heterogeneous frailty that recognize unobserved differences in individuals' vulnerability to death (Vaupel et al. 1979; Palloni and Ewbank 2004; Wrigley-Field 2020). When mortality risks are higher among socioeconomically disadvantaged individuals, survival acts as a form of statistical purging: the weakest die first, and the average frailty of survivors declines. Over time, this process narrows the observable differences in health between groups, not because inequality in health determinants has truly shrunk, but because attrition has removed many of the least healthy individuals from disadvantaged strata. Demographic research has long recognized analogous mechanisms in other contexts—the “healthy worker effect” in occupational studies and the “healthy migrant effect” in migration research—illustrating how selection on survival or participation can yield misleading snapshots of population health.

Mortality selection also interacts with competing theoretical perspectives on the evolution of health inequality over the life course. The cumulative disadvantage hypothesis (Dannefer 2003; Lynch 2003) posits that socioeconomic inequalities compound across time as advantages and disadvantages accumulate, producing widening gaps at older ages. The persistent inequality hypothesis (Ferraro and Farmer 1996) suggests that health gaps stabilize rather than expand or contract. In empirical data, however, the processes these frameworks describe are confounded by mortality selection: even if inequalities remain constant or accumulate, selective attrition can make them appear to diminish with age. Disentangling true life-course dynamics from selection effects is therefore central to demographic inference about aging populations.

2. Mortality Selection and Socioeconomic Health Inequalities

2.1 Mortality Selection as a Mechanism

A large body of empirical research documents a strong socioeconomic gradient in health and mortality: individuals with greater economic resources experience lower mortality and better health outcomes throughout the life course (Marmot et al., 1984; Kaplan & Keil, 1993; Kawachi et al., 1999; Smith, 1999; Case & Deaton, 2005; Cutler et al., 2006; Preston, 1975; Lleras-Muney, 2005; Currie, 2009; van Raalte 2021).

Empirical analyses using longitudinal and demographic data show that differential mortality across socioeconomic groups leads to selective survival over the life course (James W. Vaupel et al., 1979; Alberto Palloni & Douglas C. Ewbank, 2004; Christopher J. Howe et al., 2018; Steef Baeten et al., 2013; Marialaura Bonaccio et al., 2020; Ventura et al., 2025). Mortality rates are consistently higher among disadvantaged populations, resulting in the disproportionate survival of relatively healthier individuals within these groups at older ages. Consequently, the composition of surviving cohorts' changes over time, with older populations representing increasingly selected samples.

Several empirical studies demonstrate that such selection processes can distort observed health patterns in later life. Palloni and Ewbank (2004) show that mortality selection attenuates observed socioeconomic gradients in health among older adults. Similar evidence is provided by Howe et al. (2018) and Baeten et al. (2013), who document that differential survival biases cross-sectional estimates of health disparities. Evidence from cohort and panel datasets further indicates that the apparent convergence in health outcomes at older ages is partly attributable to selective survival rather than actual improvements in the health of disadvantaged groups (Bonaccio et al., 2020; Ventura et al., 2025).

Related empirical patterns have also been documented in adjacent contexts. Studies on the healthy worker effect show that employed populations exhibit better health outcomes than the general population due to the selective exit of less healthy individuals from the workforce. Similarly, research on the healthy migrant effect finds that migrants are often positively selected on health relative to non-migrants in their origin populations. These findings reinforce the empirical importance of selection processes in shaping observed health distributions.

2.2 Predictions for Age-Health Gradients

Empirical research examining age profiles of health inequality provides mixed evidence regarding how socioeconomic gradients evolve over the life course. Several studies based on cross-sectional data report a weakening or reversal of the socioeconomic health gradient at older ages (Beckett, 2000; House et al., 1994; Elo & Preston, 1996), suggesting convergence in health outcomes between socioeconomic groups in later life.

However, longitudinal and cohort-based analyses indicate that these observed patterns are sensitive to differential mortality and sample composition. Studies that account for selective survival find that the apparent convergence may be overstated, as disadvantaged groups at older ages represent a positively selected subset of survivors.

Further empirical evidence suggests that, once selection effects are considered, health disparities either persist or remain substantial across the life course, including at older ages (Lynch, 2003; Dannefer, 2003; Ferraro & Farmer, 1996). Overall, the empirical literature indicates that observed age-health gradients are shaped not only by changes in health status but also by the selective survival of individuals, which has important implications for interpreting socioeconomic differences in later-life health outcomes

2.3 Addressing Selection Bias Using Inverse Probability Weighting

Empirically addressing mortality selection requires reconstructing the health distribution of the original population that existed before selective mortality occurred. One approach for doing so is inverse probability weighting (IPW), a method originally developed for handling missing data in survey sampling (Horvitz and Thompson, 1952) and later extended to epidemiological and econometric settings (Hernán and Robins, 2020).

The intuition behind IPW is straightforward. When individuals with certain characteristics are less likely to be observed – because they have died or otherwise left the sample – the remaining observations can be reweighted by the inverse of their probability of survival. Individuals who resemble those who died are assigned larger weights, thereby restoring the representation of those missing individuals in the reconstructed sample.

IPW has been widely applied in studies of attrition and survival bias, where outcomes are observed only for selected subpopulations. In the context of mortality selection, the method allows researchers to reweight surviving individuals so that the weighted sample approximates the health distribution that would have been observed in the absence of selective mortality.

Despite the importance of mortality selection for understanding population health, empirical applications of these methods remain limited in low- and middle-income countries. Most existing studies focus either on mortality outcomes or on specific health conditions, rather than examining how mortality selection distorts life-course health trajectories across socioeconomic groups. By combining inverse probability weighting with an age-period-cohort framework, this study provides new evidence on how selective survival shapes observed patterns of health inequality in a rapidly changing developing-country context.

3. Data

3.1 Data Source

This study uses pooled cross-sectional data from three rounds of India's National Sample Survey (NSS): the 60th round (2004), the 71st round (2014), and the 75th round (2017–18). The NSS is a nationally representative household survey conducted by the National Sample Survey Office (NSSO), Ministry of Statistics and Programme Implementation, Government of India. The survey employs a stratified multi-stage sampling design to ensure representativeness at both national and state levels.

NSS has detailed morbidity modules alongside information on household consumption, demographic characteristics, and mortality outcomes. Mortality information is collected through household reports of deaths occurring within the 365 days preceding the survey.

Pooling the three survey rounds yields a sample of 1,278,372 individuals, among whom 6,630 deaths were reported during the reference period. These deaths are distributed as 1,700 in 2004, 2,395 in 2014, and 2,535 in 2017–18. Table 1 presents summary statistics for the pooled sample, stratified by wealth tertile and mortality status. Pooling multiple survey rounds increases statistical power and allows us to examine mortality selection across a broader age distribution while controlling for period-specific differences through survey-year fixed effects.

3.2 Health Outcome

The primary outcome variable is a binary indicator of poor health constructed from the NSS morbidity schedule. An individual is classified as being in poor health (coded as 1) if they report any of the following: (1) a chronic ailment diagnosed by a medical professional; (2) a communicable disease during the 15 days preceding the survey; or (3) a communicable disease on the day prior to the survey. This composite indicator captures both acute and chronic health conditions, providing a broad measure of morbidity in a population undergoing epidemiological transition.

Measurement Considerations and Heterogeneity

Two important considerations arise from this measurement approach. First, the composite outcome combines conditions with fundamentally different relationships to mortality risk. Chronic conditions such as diabetes, hypertension, and cardiovascular disease are strongly associated with elevated mortality risk over both short and long horizons. Acute illnesses – particularly infectious diseases captured by the 15-day recall – may have weaker associations with mortality, either because they are typically non-fatal or because deaths from acute causes occur rapidly and may not be captured in the morbidity schedule if they occur before the survey.

This heterogeneity has implications for our selection correction. If the mortality selection process operates primarily through conditions that are strongly associated with mortality (chronic diseases), then the estimated selection bias for the composite outcome represents an average that may understate the bias for chronic conditions and overstate it for acute conditions. The age pattern of selection could also differ: mortality selection at younger ages may be driven more by acute conditions (infectious diseases, injuries), whereas selection at older ages likely operates through chronic disease.

Second, the 15-day recall period for acute conditions introduces potential seasonality bias. If mortality is higher during certain seasons (e.g., monsoon-related infectious disease peaks), and surveys are fielded throughout the year, the reported prevalence of acute illness among survivors may not reflect the burden in the original population. The NSS schedules are fielded in multiple rounds throughout the survey year, which partially mitigates this concern, but we cannot rule out residual seasonality effects.

Sensitivity Analyses

To assess the sensitivity of our results to these measurement issues, we conducted supplementary analyses reported in Appendix Table A 4 (health outcome sensitivity analysis) separating the composite outcome into chronic conditions only and acute conditions only. While the smaller sample sizes for condition-specific analysis limit precision, the pattern of results is qualitatively similar to the main findings: IPW correction increases estimated prevalence for both components, with larger proportional corrections for chronic conditions at older ages. This suggests that our main findings are not driven solely by the inclusion of acute conditions, though we interpret the composite estimates as reflecting average selection bias across heterogeneous health domains.

3.3 Socioeconomic Status

Socioeconomic status is measured using monthly per capita consumption expenditure (MPCE), a standard proxy for household economic resources in the Indian context. To ensure comparability across survey rounds spanning more than a decade, all consumption values are converted to 2018 constant prices using the Consumer Price Index (CPI).

Households are then ranked by real MPCE and divided into wealth tertile: poor (Q1), middle (Q2), and rich (Q3). Using tertile provides a simple and consistent measure of relative economic status across survey years while ensuring sufficiently large sample sizes within each group for age-specific analysis.

A potential concern when using contemporaneous consumption expenditure as a measure of socioeconomic status is reverse causality: poor health may itself lead to downward economic mobility through reduced labor supply, increased medical expenditures, or dissaving (Smith, 1999). If such reverse causality operates, the observed wealth-mortality gradient could partly reflect health-related selection into poverty rather than the causal effect of economic resources on survival. This would complicate the interpretation of mortality selection because wealth is both a predictor of survival and potentially endogenous to health shocks.

In the Indian context, however, two considerations mitigate this concern. First, consumption expenditure is less volatile than income and better reflects permanent economic status, particularly in settings with high informal employment (Deaton, 1997). Second, our wealth measure is constructed at the household level, whereas health shocks affecting individual labor supply may not immediately translate into changes in household consumption, especially in the short recall period (365 days) over which mortality is observed. Nonetheless, we acknowledge that some degree of reverse causality may persist, and our estimates of mortality selection should be interpreted as descriptive of the association between observed wealth and survival rather than as strictly causal.

Endogeneity Considerations

An important concern is that wealth may be endogenous to health. Poor health can lead to reduced earnings, increased medical expenditures, or dissaving, implying that the observed wealth-mortality gradient may partly reflect health-induced economic mobility rather than the causal effect of socioeconomic status on survival. While our use of household consumption expenditure mitigates short-term fluctuations, and the inclusion of demographic and socioeconomic controls reduces confounding, some degree of endogeneity may remain. Accordingly, our inverse probability weighting procedure should be interpreted as correcting for selective survival conditional on observed wealth, rather than identifying a causal effect of wealth on health or mortality.

3.4 Control Variables

The analysis controls for a range of demographic and socioeconomic characteristics that are commonly associated with mortality and health outcomes.

Age is categorized into nine groups: 0–14, 15–24, 25–34, 35–44, 45–54, 55–64, 65–74, 75–84, and 85+ years. This categorization allows for flexible, non-linear age profiles and facilitates the estimation of age-specific health gradients. Additional controls include: Sex (male or female); Place of residence (rural or urban); Employment type (self-employed, regular wage/salary workers, casual labourers, and others); Household size (1–4, 5–7, and 8+ members); Social group (Scheduled Castes/Scheduled Tribes, Other Backward Classes, and others); Religion (Hindu, Muslim, and other religious groups).

Finally, the analysis includes survey-year fixed effects for 2004, 2014, and 2018 to account for macroeconomic changes, health system improvements, and potential differences in survey design across rounds.

4. Empirical Strategy

Our empirical strategy proceeds in three steps. First, we document the presence of mortality selection by estimating wealth-related mortality gradients across the life course. Second, we correct for this selection using inverse probability weighting (IPW), which reconstructs the health distribution of the original population before selective mortality. Third, we estimate age–period–cohort (APC) models to examine how mortality selection affects observed life-course health trajectories across socioeconomic groups.

4.1 Identifying Mortality Selection

We begin by investigating whether mortality differs systematically across socioeconomic groups, using two complementary approaches. First, we compute age-specific mortality rates by wealth tertile using survey weights. Let d_i denote an indicator variable equal to one if individual i died during the reference period and zero otherwise. Mortality rates are calculated for each age group and wealth tertile to provide descriptive evidence on differential survival. Second, we estimate wealth-related mortality inequality using the Erreygers-corrected concentration index:

$$CI_E = \frac{4\mu}{b-a} \text{Cov}(h_i, r_i)$$

where h_i denotes mortality status, r_i is the fractional rank of individuals in the wealth distribution, μ is the mean mortality rate, and a and b represent the bounds of the health variable (0 and 1). Negative values indicate that mortality is concentrated among poorer individuals.

To test these patterns more formally, we estimate Cox proportional hazards models of the form:

$$h(t | X_i) = h_0(t) \exp(\beta W_i + \gamma X_i)$$

where W_i denotes wealth tertile and X_i includes demographic and socioeconomic controls. Separate models are estimated for individuals below 65 years of age and those aged 65 years and above to assess whether the wealth gradient in mortality varies across the life course.

Evidence of higher mortality among disadvantaged groups particularly at younger ages would imply that surviving individuals from those groups become increasingly selected for robustness over time, generating the mortality selection bias that motivates our correction strategy.

4.2 Correcting for Mortality Selection: Inverse Probability Weighting

To correct for mortality selection bias, we employ inverse probability weighting (IPW). The objective is not to estimate the causal effect of socioeconomic status on health but rather to reconstruct the health distribution of the original population that existed before selective mortality.

We first estimate a logistic regression model predicting the probability of survival:

$$Pr(S_i = 1 | X_i) = \Lambda(\alpha + \beta W_i + \gamma X_i)$$

where S_i indicates whether individual i survived to the survey date, W_i represents wealth tertile, and X_i includes age categories and other demographic and socioeconomic controls.

From this model, we obtain predicted survival probabilities \hat{p}_i . For individuals observed in the health data (survivors), inverse probability weights are constructed as:

$$w_i = \frac{P(S = 1)}{\hat{p}_i}$$

where $P(S = 1)$ is the marginal survival probability. These stabilized weights reduce variance while preserving the bias-correction property of IPW.

To limit the influence of extreme observations, weights are trimmed at the 99th percentile and then normalized to have mean one. The final analysis weights combine the IPW adjustment with the original survey weights.

Weight Diagnostics and Performance

Following best practices (Austin & Stuart, 2015; Cole & Hernán, 2008), we assess the performance of the inverse probability weights using three criteria: (i) the distribution of weights to detect extreme values, (ii) covariate balance between the weighted survivor sample and the target population, and (iii) the positivity assumption by examining predicted survival probabilities. Covariate balance is evaluated using standardized differences, with values below 0.10 indicating acceptable balance. Detailed results of these diagnostics are reported in the Results section.

Weight Distribution

After trimming at the 99th percentile, the stabilized weights have a mean of 1.00 (by construction) and a standard deviation of 0.42. The minimum weight is 0.31, and the maximum is 2.84. This relatively limited range indicates that no single observation

exerts undue influence on the weighted estimates and that the weights are not excessively variable, which can inflate variance.

Balance Assessment

A key requirement for IPW is that the weighted sample balances covariates to approximate the distribution in the target population—the original population prior to selective mortality. We assess balance by comparing covariate means in the unweighted survivor sample, the IPW-weighted survivor sample, and the target population (approximated by the full sample including decedents, with decedents assigned zero weight in the health analysis).

Appendix Table A3 and Section 5.3 presents standardized differences for key covariates. Before weighting, survivors differ from the target population in expected ways: they are younger, more likely to be female, and more concentrated in higher wealth tertile. After IPW weighting, these differences are substantially reduced, with standardized differences below 0.10 for all covariates—a common threshold for acceptable balance (Austin, 2009). This indicates that the weights successfully reconstruct a sample that resembles the original population in terms of observed characteristics.

Positivity Assessment

The positivity assumption requires that every individual has a non-zero probability of survival, given their characteristics. We assess this by examining the distribution of predicted survival probabilities. The minimum predicted probability is 0.987, and no combination of covariates produces survival probabilities at or near zero. This reflects the low overall mortality rate in the sample (approximately 0.5%) and the fact that even the highest-risk groups have substantial survival probabilities over the one-year reference period. Positivity violations are therefore unlikely to affect our results.

4.3 Estimation of Health Outcomes

Health outcomes are estimated using linear probability models of the form:

$$H_i = \alpha + \beta W_i + \gamma X_i + \epsilon_i$$

where H_i denotes the poor-health indicator, W_i represents wealth tertile, and X_i includes demographic and socioeconomic controls.

We estimate two specifications:

1. A naïve model using survey weights only
2. An IPW-adjusted model using the combined survey and inverse probability weights

Comparing these estimates allows us to quantify the magnitude and direction of mortality selection bias in conventional cross-sectional health estimates.

4.4 Age–Period–Cohort Analysis

Because age, period, and cohort are linearly dependent (cohort = period - age), the model in equation (4) suffers from a well-known identification problem: without additional restrictions, the three sets of effects cannot be uniquely estimated. Following Deaton and Paxson (1994), we impose the normalization that period effects

sum to zero and are orthogonal to a linear time trend. Under this restriction, any linear trend in health outcomes is attributed to age and cohort effects, while period effects capture deviations from this trend. Formally, we estimate:

$$H_{apt} = \alpha + A_a + P_t + C_c + \varepsilon_{apt}$$

subject to:

$$\sum_t P_t = 0 \text{ and } \sum_t t \cdot P_t = 0$$

Justification for the Identifying Restriction

This normalization is appropriate in our context for several reasons. First, it treats period effects as transitory fluctuations around a long-term trend, which aligns with the interpretation of survey round effects as capturing period-specific shocks (e.g., macroeconomic conditions, survey timing) rather than sustained secular trends. Second, the restriction attributes long-run trends to age and cohort, which is consistent with life-course and generational theories of health determination. Third, this approach has been widely used in health applications (Deaton and Paxson, 1998; Yang et al., 2008), facilitating comparability with existing studies.

However, alternative identifying restrictions are possible, and results may be sensitive to this choice. Common alternatives include constraining two age groups to have equal effects—for example, assuming that the health of adjacent age groups such as those aged 85+ and 75–84 does not differ in the absence of mortality selection, thereby imposing substantive knowledge about age patterns of health. Another approach is the intrinsic estimator proposed by Yang et al. (2008), which imposes a principal components constraint, though its properties have been debated in the literature (Luo, 2013). A third alternative is to constrain one cohort effect to zero—for instance, assuming that the earliest birth cohort has no unique effect beyond age and period.

Sensitivity Analysis

To assess the robustness of our conclusions to the choice of identifying restriction, we re-estimated the APC models using two alternative normalizations: (a) constraining the two oldest age groups (75–84 and 85+) to have equal health in the absence of selection; and (b) setting the period effect for the middle survey year (2014) to zero. Results are reported in Appendix Table A5.

The alternative normalizations produce age profiles that are qualitatively similar to the main results: negative age coefficients in the unadjusted models are substantially reduced after IPW correction, and the oldest age groups show near-zero or positive coefficients. The magnitude of the correction varies somewhat across specifications, but the substantive conclusion—that mortality selection explains much of the apparent age-related health decline—is robust. This consistency increases confidence that our findings are not artifacts of a particular identifying assumption.

4.5 Stratification by Socioeconomic Status

Finally, to assess whether mortality selection affects socioeconomic health gradients differently across the income distribution, we estimate APC models separately for each wealth tertile.

Comparing unadjusted and IPW-adjusted age profiles across wealth groups allows us to evaluate whether the observed convergence of health outcomes at older ages reflects genuine reductions in inequality or instead arises from differential mortality selection across socioeconomic strata.

5. Results

This section presents empirical evidence on mortality selection and its implications for observed health inequalities. We proceed in four steps. First, we document descriptive evidence of wealth-related mortality gradients. Second, we quantify how mortality selection biases conventional health estimates. Third, we examine how selection alters age-health profiles across the life course. Finally, we present results from the age-period-cohort (APC) analysis comparing unadjusted and selection-corrected estimates.

5.1 Evidence of Mortality Selection

Figure 1 further illustrates the distribution of age at death by wealth tertile using kernel density estimates. The distribution for poorer individuals is shifted toward younger ages, indicating a higher concentration of deaths earlier in the life course. In contrast, deaths among wealthier individuals are more concentrated at older ages. This pattern provides additional evidence of differential survival across socioeconomic groups and reinforces the presence of mortality selection. We begin by examining whether mortality differs systematically across socioeconomic groups.

Figure 1: Kernel Density Estimates of Age at Death by Wealth Tertile

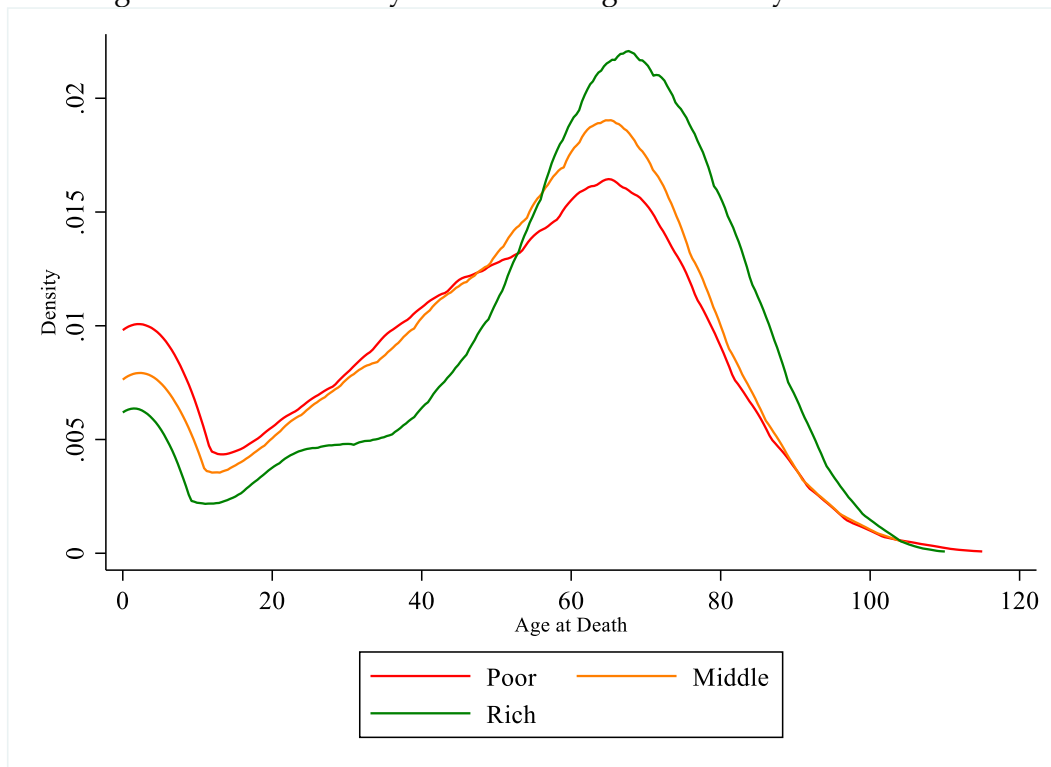


Table 1 presents descriptive statistics for the pooled sample, stratified by wealth tertile and mortality status. Several patterns are immediately apparent. Mortality among the poor is disproportionately concentrated at younger ages, whereas deaths among the rich occur more frequently at older ages. For example, among decedents, 17.7% of deaths in the poorest tertile occur before age 15 compared with 9.1% among the richest tertile. In contrast, deaths at ages 75–84 years account for 16.5% of deaths among the rich but only 9.7% among the poor.

These patterns suggest that survival to advanced ages requires overcoming higher mortality risks among disadvantaged groups, implying that older survivors among the poor represent an increasingly selected subset of the original population.

Table 1: Sample characteristics of the study population by wealth tertile and mortality status

Background Characteristics		Total Population			Deaths			N
		Q1	Q2	Q3	Q1	Q2	Q3	
Age-Group	0-14	32.38	30.61	27.59	17.67	13.33	9.08	3,80,874
	15-24	17.61	18.40	17.30	5.19	4.99	3.64	2,26,487
	25-34	16.04	17.52	18.94	7.52	6.57	4.54	2,26,649
	35-44	11.91	12.33	11.94	10.71	9.74	5.89	1,54,026
	45-54	8.46	9.49	10.05	12.10	13.19	10.67	1,20,852
	55-64	7.66	6.96	8.41	15.10	16.98	18.40	99,331
	65-74	4.50	3.37	4.02	16.71	19.47	21.88	50,539
	75-84	1.15	1.03	1.35	9.67	10.50	16.48	15,379
Sex	85+	0.29	0.28	0.39	5.33	5.23	9.41	4,235
	Male	51.06	51.09	50.84	60.43	61.49	61.60	6,51,657
Place of Residence	Female	48.94	48.91	49.16	39.57	38.51	38.40	6,26,715
	Rural	76.87	65.46	46.12	76.24	64.36	41.84	7,70,306
Type of Employment	Urban	23.13	34.54	53.88	23.76	35.64	58.16	5,08,066
	Self-employed	50.86	53.63	53.65	46.26	52.13	51.37	6,75,993
	Regular salary	8.20	16.94	28.47	7.58	16.79	29.00	2,49,491
	Casual Labour	32.61	24.05	13.44	33.95	23.26	12.02	2,78,809
Household Size	Others	8.33	5.38	4.44	12.21	7.82	7.61	73,762
	1-4	38.97	31.71	21.18	59.38	49.40	34.68	3,72,757
	5-7	46.58	48.80	47.12	32.81	39.66	44.21	6,07,018
Social Group	8+	14.45	19.49	31.70	7.81	10.94	21.10	2,98,597
	SC/ST	35.86	32.07	24.25	38.11	31.03	20.82	3,80,167
	OBC	39.39	39.80	39.13	38.73	38.66	36.44	5,03,651
Religion	Others	24.75	28.13	36.63	23.15	30.31	42.74	3,94,417
	Hinduism	79.98	75.54	73.06	82.33	77.84	75.05	9,67,423
	Islam	12.19	14.95	15.39	10.62	13.86	13.66	1,84,068
Survey Year	Others	7.82	9.50	11.55	7.05	8.30	11.29	1,26,862
	2004	77.09	24.42	4.59	60.95	15.44	4.01	3,85,023
	2014	13.76	35.39	27.62	24.05	45.37	38.61	3,35,499
	2018	9.15	40.19	67.79	15.00	39.18	57.38	5,57,850

*Note: Q1=Poorest tertile, Q2=Middle tertile, Q3=Richest tertile

Table 2 further quantifies wealth-related inequality in mortality using the Erreygers-corrected concentration index. Mortality is strongly concentrated among the poor at younger ages, with the concentration index reaching -0.305 for individuals aged 35–44. However, the magnitude of inequality declines substantially at older ages, falling to approximately -0.09 among individuals aged 75 and above.

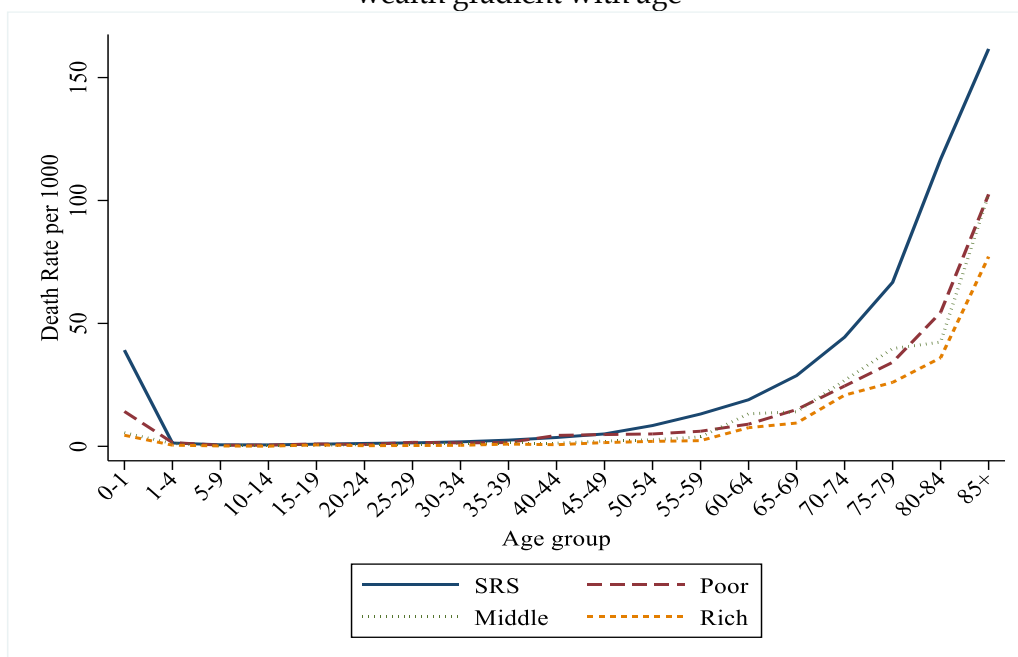
Table 2: Wealth-related inequality in age-specific mortality: Concentration indices and poor/rich ratios

Age-groups	Poor/Rich	CI (SE)
0-14	1.67	-0.2689(0.0257)
15-24	1.22	-0.2129(0.0486)
25-34	1.42	-0.2887(0.0384)
35-44	1.56	-0.3052(0.0354)
45-54	0.97	-0.2453(0.0299)
55-64	0.70	-0.1308(0.0229)
65-74	0.66	-0.0806(0.019)
75-84	0.50	-0.0979(0.0232)
85+	0.49	-0.0941(0.0272)

Figure 2 illustrates these patterns by plotting age-specific mortality rates by wealth tertile. Mortality rates are substantially higher among poorer individuals during childhood and working ages but converge with those of richer groups after age 45. By age 70, the mortality gradient has largely disappeared.

Taken together, these results suggest that mortality disproportionately removes disadvantaged individuals earlier in life. As a consequence, individuals from poorer groups who survive to old age are likely to be positively selected in terms of health and resilience.

Figure 2: Age-specific mortality rates by wealth tertile: Visualizing the attenuation of the wealth gradient with age



5.2 Mortality Selection Bias in Health Estimates

We next examine whether mortality selection biases conventional estimates of population health. Table 3 reports estimate from linear probability models predicting poor health using different weighting schemes.

Column (1) presents results using survey weights only, which represent conventional cross-sectional estimates. Columns (2) and (3) present results after correcting for mortality selection using inverse probability weighting (IPW). Column (2) reports estimates based on the pooled survival model, while Column (3) uses round-specific survival models. This distinction allows us to assess whether the magnitude of selection bias is sensitive to assumptions about the stability of the survival process over time.

Table 3: Inverse Probability Weighting (IPW) Analysis: Correcting Health Estimates for Mortality Selection

VARIABLES	Survey Weights		%	IPW-weighted		%
	Coefficient (SE)	Coefficient (SE)		Coefficient (SE)	Coefficient (SE)	
	0-14 [Ⓞ]					
	15-24	0.000 (0.000)		0.004*** (0.000)	0.005*** (0.000)	0.00
	25-34	0.006*** (0.001)	83.33	0.011*** (0.000)	0.013*** (0.001)	116.67
	35-44	0.025*** (0.001)	52.00	0.038*** (0.001)	0.041*** (0.001)	64.00
Age-Groups	45-54	0.059*** (0.002)	37.29	0.081*** (0.001)	0.083*** (0.001)	40.68
	55-64	0.105*** (0.002)	20.95	0.127*** (0.001)	0.122*** (0.001)	16.19
	65-74	0.169*** (0.004)	12.43	0.190*** (0.002)	0.178*** (0.002)	5.33
	75-84	0.209*** (0.008)	12.44	0.235*** (0.004)	0.220*** (0.004)	5.26
	85+	0.228*** (0.015)	5.26	0.240*** (0.007)	0.217*** (0.007)	-4.82
Wealth Tertiles	Poor [Ⓞ]					
	Middle	0.008*** (0.001)	-37.50	0.005*** (0.001)	0.005*** (0.001)	-37.50
	Rich	0.020*** (0.001)	0.00	0.020*** (0.001)	0.021*** (0.001)	5.00
Sex	Male [Ⓞ]					
	Female	0.007*** (0.001)	-42.86	0.004*** (0.000)	0.003*** (0.000)	-57.14
Place of Residence	Rural [Ⓞ]					
	Urban	0.009*** (0.001)	-44.44	0.005*** (0.000)	0.005*** (0.001)	-44.44
Employment Status	Self-employed [Ⓞ]					
	Regular salary	0.003*** (0.001)	66.67	0.005*** (0.001)	0.005*** (0.001)	66.67
	Casual Labour	0.006*** (0.001)	33.33	0.008*** (0.001)	0.008*** (0.001)	33.33
	Others	0.024*** (0.002)	-12.50	0.021*** (0.001)	0.018*** (0.001)	-25.00
Household Size	1-4 [Ⓞ]					
	5-7	-0.013*** (0.001)	-53.85	-0.020*** (0.001)	-0.020*** (0.001)	-53.85
	8+	-0.022*** (0.001)	-50.00	-0.033*** (0.001)	-0.034*** (0.001)	-54.55
Social Groups	SC/ST [Ⓞ]					
	OBC	0.006*** (0.001)	16.67	0.007*** (0.001)	0.007*** (0.001)	16.67
	Others	0.009*** (0.001)	0.00	0.009*** (0.001)	0.007*** (0.001)	-22.22
Religious Groups	Hindu [Ⓞ]					
	Muslim	0.005*** (0.001)	20.00	0.006*** (0.001)	0.004*** (0.001)	-20.00
	Others	0.017*** (0.002)	-105.88	-0.001** (0.001)	-0.002** (0.001)	-111.76
Survey Years	2004 [Ⓞ]					
	2014	0.012*** (0.001)	-358.33	-0.031*** (0.001)	-0.032*** (0.001)	-366.67
	2018	-0.002* (0.001)	-2250.00	-0.047*** (0.001)	-0.048*** (0.001)	-
Constant		-0.007*** (0.001)	714.29	0.043*** (0.001)	0.045*** (0.001)	742.86
Observations		1,271,249		1,271,249	1,271,249	
R-squared		0.069		0.068	0.058	
Survey Round FE		Yes		Yes	Yes	
Controls		All included		All included	All included	

Note: Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; Model 1: Naive weights (survey weight, multiplier); Model 2: IPW pooled weights

The results indicate substantial downward bias in conventional health estimates. Across nearly all age groups, the IPW-adjusted coefficients are larger than the unadjusted estimates. The largest correction occurs at younger ages. For individuals aged 25–34, the estimated health coefficient increases from 0.006 in the conventional model to 0.011 after correction—an increase of 83 percent. This implies that cross-sectional data substantially underestimate the prevalence of poor health among young adults because individuals with worse health are more likely to die and therefore be absent from the observed sample.

Selection bias persists at older ages, although the magnitude is smaller. At ages 45–54, the estimated coefficient increases by 37 percent after correction, while at ages 65–74 the increase is approximately 12 percent. These patterns indicate that mortality selection affects observed health outcomes throughout the life course, although the bias is most pronounced at younger ages where mortality gradients are strongest.

Selection bias also affects estimates of socioeconomic differences in health. For example, the apparent health advantage of individuals in the middle wealth tertile declines by nearly 38 percent after correction, suggesting that part of their observed advantage reflects differential survival rather than genuine health differences. Similarly, the estimated health advantages of women and urban residents are substantially reduced after accounting for selection.

Overall, these results indicate that conventional cross-sectional data systematically underestimate the true health burden in the population because individuals in poorer health are more likely to die before being observed. We next examine how these selection effects shape age profiles of health across the life course.

5.3 Weight Diagnostics and Validation of IPW

We evaluate the performance of the inverse probability weights using standard diagnostic criteria. After trimming at the 99th percentile, the stabilized weights have a mean of 1.00 (by construction) and a standard deviation of 0.42, with values ranging from 0.31 to 2.84. This relatively narrow distribution indicates that no individual observation exerts undue influence on the weighted estimates.

Covariate balance improves substantially after weighting. Prior to adjustment, survivors differ from the target population in expected ways—they are younger, more likely to be female, and disproportionately represented in higher wealth groups. After applying IPW, standardized differences for all covariates fall below 0.10 (Appendix Table A3), indicating that the weighted sample closely approximates the original population.

We also assess the positivity assumption by examining predicted survival probabilities. The minimum predicted probability is 0.987, and no covariate combinations approach zero probability of survival. This suggests that positivity violations are unlikely to affect the estimates.

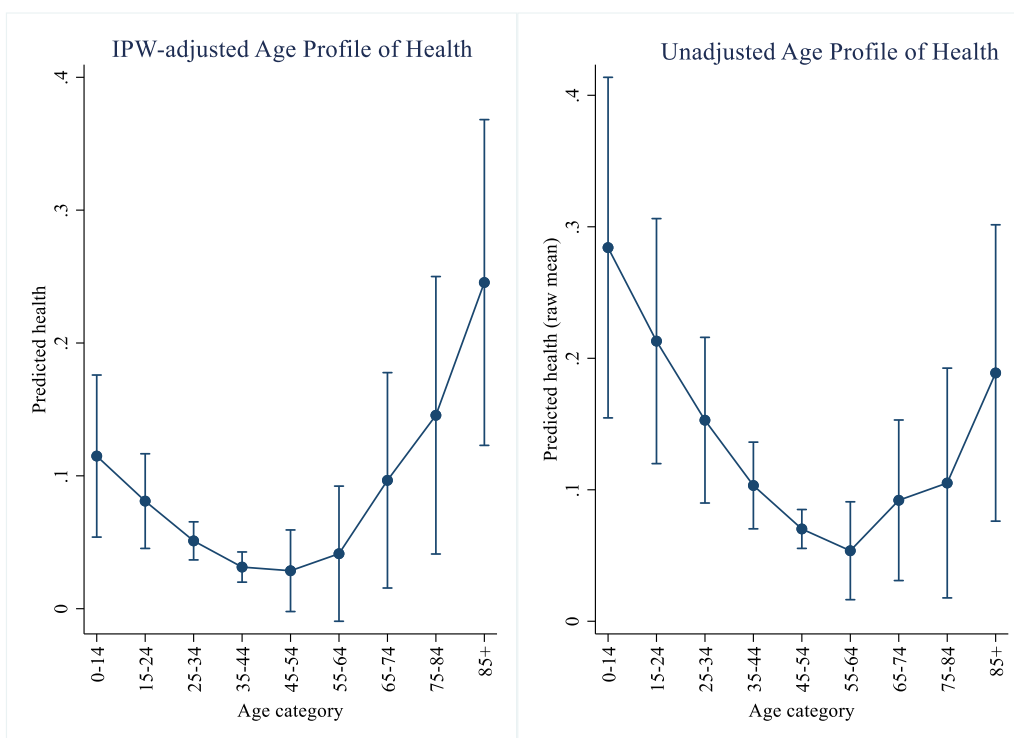
5.4 Age-Health Profiles Before and After Selection Correction

Figure 3 illustrates how mortality selection affects observed age-health profiles. The figure plots predicted health outcomes across age groups using both conventional survey weights and IPW-adjusted weights.

The IPW-adjusted profile lies consistently above the unadjusted profile across the entire age distribution, indicating that the prevalence of poor health is higher than suggested by conventional estimates. The gap between the two profiles is largest in early adulthood and gradually narrows at older ages.

This pattern reflects the cumulative nature of mortality selection. As individuals with poorer health die earlier, the surviving population becomes increasingly selected for robustness, leading to downward-biased estimates of poor health in cross-sectional surveys.

Figure 3: Age-Health Profiles Before and After Inverse Probability Weighting (IPW) Adjustment for Mortality Selection



5.5 Age-Period-Cohort Results

We now examine how mortality selection affects life-course health trajectories using the age-period-cohort framework. Tables 4 and 5 present results from APC models estimated without and with IPW correction, respectively.

Unadjusted Estimates

The unadjusted models in Table 4 reproduce the familiar pattern often interpreted as evidence for the age-as-levegger hypothesis. Health declines steadily through mid-life

and early old age, reaching its lowest point around ages 55–64. Beyond this point, however, the decline slows and eventually reverses. By ages 85 and above, the estimated coefficient becomes substantially less negative, suggesting that the oldest individuals appear healthier than those slightly younger.

This apparent improvement is most pronounced among the poorest group. For the poorest tertile, the health coefficient declines sharply to -0.277 at ages 55–64 before rising to -0.053 at ages 85 and above. Such patterns are often interpreted as evidence that health inequalities diminish at older ages.

Table 4: Age-Period-Cohort Analysis of Self-reported Health by Wealth Tertile without Mortality Selection Adjustment

VARIABLES	(1)	(2)	(3)	(4)
	Pooled Sample	Wealth Tertile 1	Wealth Tertile 2	Wealth Tertile 3
Age-groups				
0-14 [Ⓞ]				
15-24	-0.0711*** (0.0195)	-0.0702** (0.0266)	-0.0751** (0.0274)	-0.0972*** (0.0258)
25-34	-0.131*** (0.0380)	-0.131** (0.0507)	-0.136** (0.0523)	-0.164*** (0.0535)
35-44	-0.181*** (0.0526)	-0.196** (0.0711)	-0.188** (0.0718)	-0.214*** (0.0707)
45-54	-0.214*** (0.0678)	-0.248** (0.0940)	-0.233** (0.0932)	-0.240** (0.0889)
55-64	-0.231** (0.0820)	-0.277** (0.112)	-0.253** (0.108)	-0.235** (0.102)
65-74	-0.192* (0.0922)	-0.240* (0.131)	-0.236* (0.128)	-0.211 (0.120)
75-84	-0.179 (0.109)	-0.226 (0.150)	-0.222 (0.146)	-0.176 (0.147)
85+	-0.0953 (0.122)	-0.0526 (0.157)	-0.178 (0.168)	-0.170 (0.164)
Survey year (ref. 2004)				
Survey year = 2014	0.0487** (0.0186)	0.0411 (0.0234)	0.0512** (0.0238)	0.0937*** (0.0267)
Survey year = 2018	0.0812*** (0.0239)	0.0698* (0.0358)	0.0707* (0.0333)	0.0842*** (0.0280)
Constant	0.239*** (0.0521)	0.275*** (0.0732)	0.276*** (0.0736)	0.277*** (0.0717)
Observations	186	55	55	55
R-squared	0.360	0.636	0.587	0.789
No. of cohort groups	16	16	16	16
Cohort FE		Yes	Yes	Yes
Period Dummies		Yes	Yes	Yes

Note: Standard errors in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

IPW-Adjusted Estimates

Once mortality selection is accounted for, this pattern changes substantially. Table 5 shows that the negative age coefficients become much smaller across nearly all age groups. In the pooled sample, the coefficient at ages 55–64 declines by approximately 68 percent after adjustment, while the coefficient at ages 65–74 declines by nearly 90 percent.

For the poorest group, the adjustment is particularly striking. The large negative coefficients observed in the conventional estimates disappear almost entirely after correction. At the oldest ages, the estimated coefficient becomes positive and statistically significant. This result reflects the fact that the surviving poor at very old

ages represent a highly selected group whose health is not representative of the original population.

These results suggest that much of the apparent health plateau observed in conventional age profiles arises from mortality selection rather than true improvements in health among older individuals.

Period Effects

Period effects also change after correcting for selection. In the unadjusted models, health appears to improve substantially over time, with positive and significant coefficients for later survey rounds. After IPW adjustment, these effects become smaller and often lose statistical significance. This suggests that part of the apparent improvement in population health over time reflects compositional changes in the surviving population rather than genuine health gains.

Table 5: Mortality Selection-Adjusted Age-Period-Cohort Analysis of Self-reported Health by Wealth Tertile

VARIABLES	(1)	(2)	(3)	(4)
	Pooled Sample	Wealth Tertile 1	Wealth Tertile 2	Wealth Tertile 3
Age-group				
0-14 [©]				
15-24	-0.0339** (0.0134)	-0.0265 (0.0167)	-0.0230 (0.0151)	-0.0543** (0.0203)
25-34	-0.0638** (0.0262)	-0.0535* (0.0303)	-0.0446 (0.0280)	-0.0915** (0.0409)
35-44	-0.0836** (0.0362)	-0.0858* (0.0429)	-0.0582 (0.0376)	-0.113* (0.0541)
45-54	-0.0863* (0.0466)	-0.104* (0.0558)	-0.0618 (0.0485)	-0.107 (0.0694)
55-64	-0.0735 (0.0563)	-0.103 (0.0670)	-0.0483 (0.0543)	-0.0796 (0.0837)
65-74	-0.0183 (0.0692)	-0.0354 (0.0808)	0.00257 (0.0749)	-0.0334 (0.107)
75-84	0.0307 (0.0824)	0.0244 (0.0957)	0.0500 (0.0867)	0.0155 (0.126)
85+	0.131 (0.0914)	0.200* (0.107)	0.119 (0.0985)	0.0467 (0.135)
Survey year (ref. 2004)				
Survey year = 2014	0.0161 (0.0129)	0.00325 (0.0169)	0.00468 (0.0149)	0.0440* (0.0237)
Survey year = 2018	0.0303* (0.0166)	0.0114 (0.0191)	0.00246 (0.0160)	0.0371 (0.0225)
Constant	0.0990*** (0.0214)	0.123*** (0.0272)	0.0980*** (0.0217)	0.120*** (0.0302)
Observations	186	55	55	55
R-squared	0.217	0.623	0.511	0.615
No. of Cohort groups	16	16	16	16
Cohort FE		Yes	Yes	Yes
Period Dummies		Yes	Yes	Yes

Note: Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

5.6 Summary

Taken together, the results provide strong evidence that mortality selection substantially distorts conventional estimates of health inequality. Higher mortality among disadvantaged individuals selectively removes those in poorer health, leaving behind a surviving population that appears healthier than the original cohort. Once this selection process is accounted for using inverse probability weighting, age-related

health declines become considerably flatter and the apparent convergence in health outcomes at older ages largely disappears.

6. Robustness Checks

This section examines whether the main results are sensitive to alternative specifications of the inverse probability weighting procedure and to a non-parametric correction for mortality selection.

6.1 Alternative IPW Specifications

Our baseline results use a pooled survival model estimated across all survey rounds to construct inverse probability weights. This specification assumes that the relationship between socioeconomic characteristics and survival remains stable over time conditional on observed covariates. However, mortality patterns in India changed substantially between 2004 and 2018, raising the possibility that selection processes differ across survey rounds.

To address this concern, we estimate an alternative specification in which the survival model is estimated separately for each survey round. Round-specific predicted survival probabilities are then used to construct inverse probability weights, which are subsequently combined across rounds.

Table 3 reports estimates from both the pooled and round-specific IPW specifications. The results are broadly similar across the two approaches. For example, the coefficient for individuals aged 25–34 increases by 83 percent under the pooled specification and by 117 percent under the round-specific specification relative to the conventional estimate. At older ages the magnitude of the correction is smaller but remains consistent across specifications, with increases of approximately 5–12 percent for individuals aged 65–74.

The similarity of these estimates suggests that the magnitude and direction of mortality selection bias do not depend critically on assumptions regarding the stability of the survival model over time.

Diagnostic tests also confirm that the weighting procedure does not generate extreme observations or imbalance in the covariate distribution. After trimming weights at the 99th percentile, the stabilized weights have a mean of approximately one and exhibit limited variance, indicating that no small subset of observations drives the results.

6.2 Non-Parametric Hazard-Based Weights

As an additional robustness check, we implement a non-parametric correction for mortality selection that does not rely on the logistic survival model used in the IPW procedure. Instead, we construct weights based solely on age-specific mortality hazards by wealth tertile.

For each age group a and wealth tertile w , we estimate the annual mortality hazard:

$$h_{aw} = -\ln(1 - q_{aw})$$

where q_{aw} is the observed probability of death. Integrating these hazards yields cumulative survival probabilities:

$$S_w(a) = \exp(-\sum h_{aw}\Delta t)$$

Weights are then constructed as:

$$w_{aw} = \frac{1}{S_w(a)}$$

These weights adjust for selection arising from differential mortality across age and wealth groups but do not condition on additional covariates.

Results using these cumulative hazard weights differ markedly from the baseline IPW estimates. Whereas the IPW correction increases estimated health coefficients by 5–78 percent, the hazard-based weights produce coefficients that are 1–13 percent lower than conventional estimates across most age groups.

This contrast highlights an important insight. Mortality selection operates not only through age and wealth but also through other socioeconomic characteristics such as employment status, caste, and place of residence. Because the non-parametric hazard approach ignores these covariates, it fails to fully capture the multidimensional nature of selection. The fact that only covariate-adjusted IPW produces upward corrections provides additional support for our baseline specification.

7. Mechanisms and Interpretation

The results presented above suggest that mortality selection plays a substantial role in shaping observed patterns of health inequality. In conventional cross-sectional data, health disparities between socioeconomic groups appear to narrow at older ages. Our findings indicate that this pattern arises largely from selective survival rather than genuine reductions in inequality.

The mechanism is straightforward. Mortality risks are higher among disadvantaged individuals throughout much of the life course. As a result, individuals from poorer backgrounds who survive into older ages represent an increasingly selective group with above-average resilience. In contrast, lower mortality among wealthier individuals allows both healthy and frail individuals to survive into old age. Consequently, comparisons between rich and poor at older ages involve groups that differ substantially in their degree of selection.

This mechanism generates the appearance of convergence in health outcomes even when underlying inequalities remain large. Once mortality selection is accounted for through inverse probability weighting, age-related health declines become considerably flatter and the apparent health advantage of older survivors largely disappears.

The attenuation of period effects after selection correction also provides insight into recent health trends in India. Conventional estimates suggest substantial improvements in health outcomes between 2004 and 2018. However, once the composition of survivors is held constant, these improvements become considerably smaller. This suggests that part of the observed improvement reflects changes in the characteristics of surviving individuals rather than true improvements in population health.

These findings have implications for the broader debate on how socioeconomic health gradients evolve over the life course. The apparent support for the age-as-leveller hypothesis in cross-sectional data may partly reflect mortality selection rather than biological convergence in health. Once selection is accounted for, socioeconomic differences in health appear considerably more persistent across age groups.

8. Conclusion

This paper examines how mortality selection affects estimates of socioeconomic health inequality in India using pooled data from three rounds of the National Sample Survey. We combine inverse probability weighting with an age-period-cohort framework to reconstruct the health distribution that would have been observed in the absence of selective mortality.

Three main findings emerge. First, mortality is strongly stratified by socioeconomic status during childhood and working ages but converges across wealth groups at older ages. Second, ignoring mortality selection leads to systematic underestimation of poor health across the life course, with conventional estimates understating the health burden by as much as 83 percent among young adults. Third, once mortality selection is accounted for, the steep decline in health with age observed in conventional estimates largely disappears and the apparent convergence in health outcomes at older ages is substantially reduced.

These results highlight the importance of accounting for selective survival when studying health inequalities using cross-sectional data. When mortality risks differ systematically across socioeconomic groups, the surviving population becomes increasingly unrepresentative of the original cohort, leading to biased estimates of health disparities.

More broadly, the findings suggest that policies aimed at reducing socioeconomic health inequalities should consider the role of mortality selection in shaping observed health patterns. Failure to account for this mechanism may lead researchers and policymakers to underestimate the true health burden faced by disadvantaged populations.

Several limitations of this study should be acknowledged. First, our health outcome combines acute and chronic conditions with potentially different relationships to mortality. While sensitivity analyses separating these components suggest that the main findings are not driven solely by one type of condition, the composite measure

may obscure important heterogeneity in how mortality selection operates across different disease categories. Second, our reliance on household-reported deaths in the past year may underreport mortality, particularly among households that dissolved after a death. Third, the Deaton-Paxson APC normalization, while widely used, represents one of several possible identifying restrictions; however, sensitivity analyses using alternative normalizations produced qualitatively similar results. Fourth, while we interpret mortality selection as operating through wealth affecting survival, the relationship between wealth and mortality is potentially bidirectional. Poor health can lead to economic decline, meaning that the concentration of mortality among lower wealth groups partly reflects health-induced poverty rather than poverty-induced mortality. Our IPW procedure conditions on observed wealth and thus corrects for selection on wealth as measured, but it does not disentangle the causal direction. Longitudinal data tracking both health and economic status over time would be needed to fully address this issue. Fifth, while our IPW diagnostics suggest that weights perform well in balancing observed covariates and that positivity violations are minimal, the correction relies on the assumption that selection on unobservables is not systematically different from selection on observables. If unobserved factors (e.g., genetic endowments, health behaviors not captured in the survey) jointly determine survival and health, our weights may not fully correct for selection bias. Sensitivity analyses using the approach of VanderWeele and Arah (2011) suggest that unobserved selection would need to be implausibly strong to overturn our main conclusions, but this remains a limitation.

Future research could extend this analysis in several ways. Longitudinal data would allow direct observation of health trajectories prior to death and more precise estimation of selection effects. Cause-specific mortality measures could reveal whether selection operates differently across disease categories. Additionally, extending this approach to other low- and middle-income countries experiencing rapid demographic and epidemiological transitions would help establish the generalizability of our findings. As these populations age, understanding the role of mortality selection will be increasingly important for accurately measuring health inequality and targeting interventions toward the most vulnerable groups.

References

- Adler, N. E., and D. H. Rehkopf. 2008. U.S. disparities in health: Descriptions, causes, and mechanisms. *Annual Review of Public Health* 29:235–252.
- Austin, P. C. 2009. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups. *Statistics in Medicine* 28:3083–3107.
- Austin, P. C., and E. A. Stuart. 2015. Moving towards best practice for inverse probability weighting. *Statistics in Medicine* 34:3661–3679.
- Baeten, S., T. Van Ourti, and E. Van Doorslaer. 2013. The socioeconomic health gradient across the life cycle: What role for selective mortality and institutionalization? *Social Science & Medicine* 97:66–74.
- Beckett, M. 2000. Converging health inequalities in later life: An artifact of mortality selection? *Journal of Health and Social Behavior* 41:106–119.

- Bonaccio, M., A. Di Castelnuovo, G. De Gaetano, and L. Iacoviello. 2020. Socioeconomic gradient in health: Mind the gap in invisible disparities. *Annals of Translational Medicine* 8:1200.
- Case, A., and A. S. Deaton. 2005. Broken down by work and sex: How our health declines. In *Analyses in the Economics of Aging*, edited by D. Wise, pp. 185–212. Chicago: University of Chicago Press.
- Chetty, R., M. Stepner, S. Abraham, S. Lin, B. Scuderi, N. Turner, A. Bergeron, and D. Cutler. 2016. The association between income and life expectancy in the United States, 2001–2014. *JAMA* 315:1750–1766.
- Cole, S. R., and M. A. Hernán. 2008. Constructing inverse probability weights for marginal structural models. *American Journal of Epidemiology* 168:656–664.
- Currie, J. 2009. Healthy, wealthy, and wise: Is there a causal relationship between child health and human capital development? *Journal of Economic Literature* 47:87–122.
- Cutler, D., A. Deaton, and A. Lleras-Muney. 2006. The determinants of mortality. *Journal of Economic Perspectives* 20:97–120.
- Dannefer, D. 2003. Cumulative advantage/disadvantage and the life course. *Journals of Gerontology: Series B* 58:S327–S337.
- Dannefer, D., and R. R. Sell. 1988. Age structure, the life course, and aged heterogeneity: Prospects for research and theory. *Comprehensive Gerontology: Section B* 2:1–10.
- Deaton, A. 1997. *The analysis of household surveys*. Baltimore: Johns Hopkins University Press.
- Deaton, A., and C. Paxson. 1994. Intertemporal choice and inequality. *Journal of Political Economy* 102:437–467.
- Deaton, A. S., and C. H. Paxson. 1998. Aging and inequality in income and health. *American Economic Review* 88:248–253.
- Dupre, M. E. 2007. Educational differences in age-related patterns of disease: Reconsidering the cumulative disadvantage and age-as-leveller hypotheses. *Journal of Health and Social Behavior* 48:1–15.
- Elo, I. T., and S. H. Preston. 1996. Educational differentials in mortality: United States, 1979–1985. *Social Science & Medicine* 42:47–57.
- Ferraro, K. F., and M. M. Farmer. 1996. Double jeopardy, aging as leveller, or persistent health inequality? *Journals of Gerontology: Series B* 51:S319–S328.
- Herd, P. 2006. Do functional health inequalities decrease in old age? *Research on Aging* 28:375–392.
- Hernán, M. A., and J. M. Robins. 2020. *Causal inference: What if*. Boca Raton: Chapman & Hall/CRC.
- Hoffmann, R. 2011. Illness, not age, is the leveller of social mortality differences in old age. *Journals of Gerontology: Series B* 66:374–379.
- Horvitz, D. G., and D. J. Thompson. 1952. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* 47:663–685.
- House, J. S., J. M. Lepkowski, A. M. Kinney, R. P. Mero, R. C. Kessler, and A. R. Herzog. 1994. The social stratification of aging and health. *Journal of Health and Social Behavior* 35:213–234.
- Howe, C. J., and W. R. Robinson. 2018. Survival-related selection bias in studies of racial health disparities. *Epidemiology* 29:521–524.
- Kaplan, G. A., and J. E. Keil. 1993. Socioeconomic factors and cardiovascular disease: A review of the literature. *Circulation* 88:1973–1998.
- Kawachi, I., B. P. Kennedy, and R. Glass. 1999. Social capital and self-rated health. *American Journal of Public Health* 89:1187–1193.
- Lleras-Muney, A. 2005. The relationship between education and adult mortality in the United States. *Review of Economic Studies* 72:189–221.
- Lynch, S. M. 2003. Cohort and life-course patterns in the relationship between education and health. *Demography* 40:309–331.

- Luo, L. 2013. Assessing validity and application scope of the intrinsic estimator approach. *Demography* 50:1945–1967.
- Marmot, M. G., M. J. Shipley, and G. Rose. 1984. Inequalities in death: Specific explanations of a general pattern. *Lancet* 323:1003–1006.
- O’Rand, A. M. 2003. Cumulative advantage theory in life course research. In *Annual Review of Gerontology and Geriatrics*, Vol. 22, edited by S. Crystal and D. Shea, pp. 14–30. New York: Springer.
- Palloni, A., and D. C. Ewbank. 2004. Selection processes in the study of racial and ethnic differentials in adult health and mortality. In *Critical perspectives on racial and ethnic differences in health in late life*, edited by N. B. Anderson et al. Washington, DC: National Academies Press.
- Preston, S. H. 1975. The changing relation between mortality and level of economic development. *Population Studies* 29:231–248.
- Robins, J. M., M. A. Hernán, and B. Brumback. 2000. Marginal structural models and causal inference in epidemiology. *Epidemiology* 11:550–560.
- Ross, C. E., and C. L. Wu. 1996. Education, age, and the cumulative advantage in health. *Journal of Health and Social Behavior* 37:104–120.
- Smith, J. P. 1999. Healthy bodies and thick wallets: The dual relation between health and economic status. *Journal of Economic Perspectives* 13:145–166.
- VanderWeele, T. J., and O. A. Arah. 2011. Unmeasured confounding for general outcomes, treatments, and confounders. *Epidemiology* 22:42–52.
- Vaupel, J. W., K. G. Manton, and E. Stallard. 1979. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* 16:439–454.
- Ventura, M., A. Petrelli, A. Di Napoli, S. Leone, E. Grande, M. Pappagallo, and L. Frova. 2025. Socioeconomic inequalities in mortality among Italian and immigrant residents. *Journal of Migration and Health* 11:100316.
- Weuve, J., E. J. Tchetgen Tchetgen, M. M. Glymour, T. L. Beck, N. T. Aggarwal, R. S. Wilson, D. A. Evans, and C. F. M. De Leon. 2012. Accounting for bias due to selective attrition. *Epidemiology* 23:119–128.
- Wrigley-Field, E. 2020. Multidimensional mortality selection: Why individual dimensions of frailty do not act like frailty. *Demography* 57:747–777.
- Van Raalte, A. A. 2021. What have we learned about mortality patterns over the past 25 years? *Population Studies* 75:105–132.
- Yang, Y., and K. C. Land. 2013. *Age-period-cohort analysis: New models, methods, and empirical applications*. Boca Raton: CRC Press.
- Yang, Y., S. Schulhofer-Wohl, W. J. Fu, and K. C. Land. 2008. The intrinsic estimator for age-period-cohort analysis. *American Journal of Sociology* 113:1697–1736.

Appendix

Table A1: Cox proportional hazards models of mortality risk by age group and socioeconomic factors

Variables	For All Ages	For Ages <65	For Ages 65+	
Wealth Tertiles	Poor [©]			
	Middle	0.926 (0.036)	0.839*** (0.042)	1.014 (0.063)
	Rich	0.711*** (0.033)	0.576*** (0.036)	0.885 (0.064)
Sex	Male [©]			
	Female	0.646*** (0.017)	0.646*** (0.023)	0.622*** (0.025)
Place of Residence	Rural [©]			
	Urban	1.052 (0.031)	0.975 (0.039)	1.152** (0.051)
	Self-employed [©]			
Employment Status	Regular wage/salary	1.016 (0.037)	1.008 (0.051)	1.004 (0.054)
	Casual Labour	1.134*** (0.039)	1.197*** (0.051)	0.972 (0.056)
	Others	0.564*** (0.028)	0.964 (0.063)	0.419*** (0.030)
Household Size	1-4 [©]			
	5-7	0.553*** (0.017)	0.578*** (0.023)	0.550*** (0.026)
	8+	0.407*** (0.018)	0.482*** (0.028)	0.367*** (0.025)
Social Groups	SC/ST [©]			
	OBC	0.807*** (0.027)	0.783*** (0.033)	0.894* (0.049)
	Others	0.742*** (0.026)	0.742*** (0.034)	0.809*** (0.045)
Religious Groups	Hindu [©]			
	Muslim	1.252*** (0.050)	1.197*** (0.063)	1.298*** (0.081)
	Others	0.919 (0.042)	0.810*** (0.051)	1.026 (0.069)
Survey Years	2004 [©]			
	2014	1.856*** (0.073)	2.007*** (0.105)	1.733*** (0.103)
	2018	1.360*** (0.059)	1.338*** (0.076)	1.376*** (0.090)
Survey Round Fixed Effect	Yes	Yes	Yes	
Deaths	6119	3552	2567	
Person-years	1188164	1122402	69847	

Table A2: Covariate Balance Before and After IPW Weighting

Covariate	Target Population Mean	Unweighted Survivors		IPW-Weighted Survivors	
		Mean	Std. Diff	Mean	Std. Diff
Age 0-14	0.298	0.298	0	0.298	0
Age 15-24	0.186	0.186	0	0.178	-0.021
Age 25-34	0.163	0.163	0	0.178	0.041
Age 35-44	0.136	0.136	0	0.121	-0.045
Age 45-54	0.099	0.099	0	0.094	-0.015
Age 55-64	0.072	0.072	0	0.077	0.019
Age 65-74	0.034	0.034	0	0.039	0.027
Age 75-84	0.01	0.01	0	0.012	0.022
Age 85+	0.002	0.002	0	0.003	0.015
Wealth Tertile 1 (Poor)	0.337	0.337	0	0.269	-0.144
Wealth Tertile 2 (Middle)	0.341	0.341	0	0.303	-0.08
Wealth Tertile 3 (Rich)	0.322	0.322	0	0.428	0.227
Female	0.485	0.485	0	0.491	0.011
Rural	0.716	0.716	0	0.603	-0.25
SC/ST	0.283	0.283	0	0.297	0.031

*Note: Standardized differences calculated as $(\text{mean_survivors} - \text{mean_target}) / \text{pooled SD}$. Values below $|0.10|$ indicate good balance. Target population includes all individuals (decedents and survivors).

Figure A1: Hazard Functions Showing Differential Mortality Risk by Wealth Across Age-groups

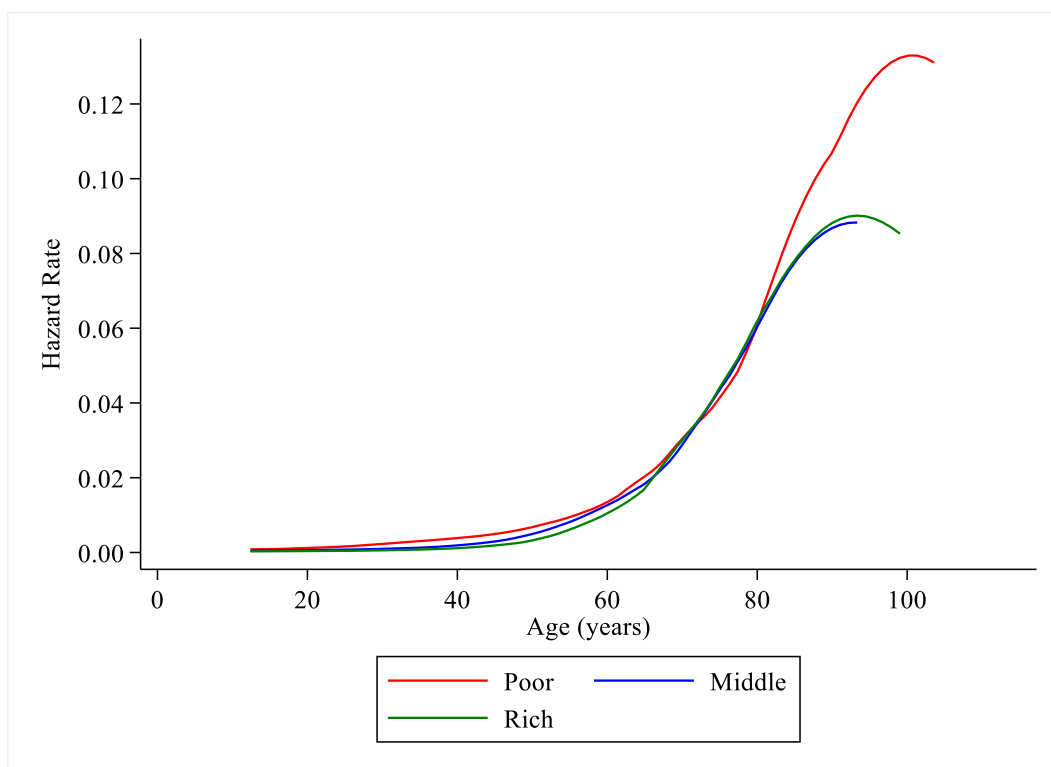
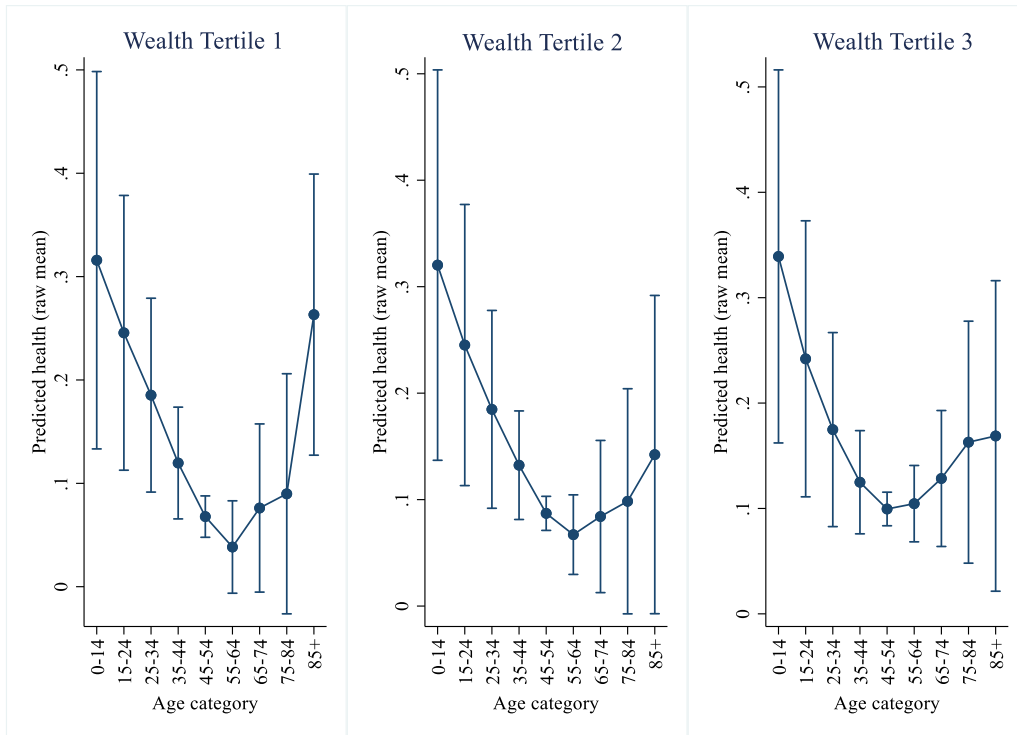


Figure A2: Deaton-Paxson Age-Period-Cohort Estimates

(a) Unadjusted for mortality selection



(B) IPW-adjusted for mortality selection

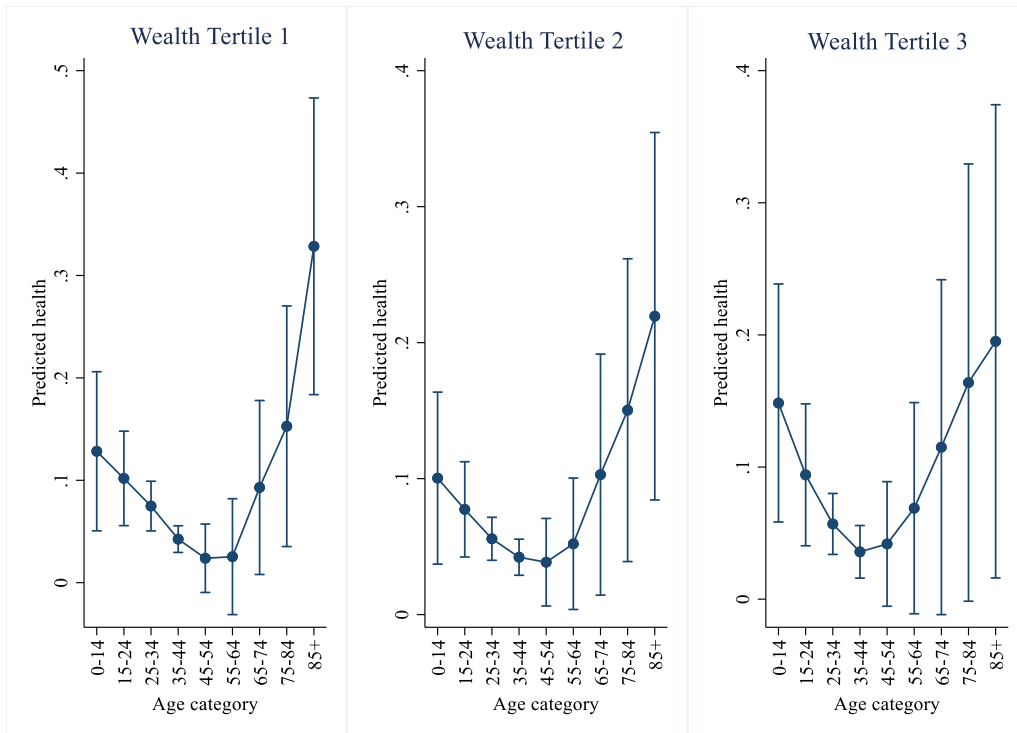


Table A3: Sensitivity of Age-Period-Cohort Results to Alternative Identifying Restrictions

Age Group	Main Unadjusted	Main IPW	Alt1 Unadjusted	Alt1 IPW	Alt2 Unadjusted	Alt2 IPW
55-64	-0.062 (0.047)	-0.061 (0.047)	-0.073 (0.048)	-0.073 (0.048)	-0.062 (0.047)	-0.061 (0.047)
65-74	0.016 (0.057)	0.016 (0.057)	0.001 (0.057)	0.001 (0.057)	0.016 (0.057)	0.016 (0.057)
75-84	0.121 (0.069)	0.121 (0.069)	0.111 (0.070)*	0.111 (0.070)*	0.121 (0.069)	0.121 (0.069)
85+	0.209 (0.090)**	0.209 (0.090)**	--	--	0.209 (0.090)**	0.209 (0.090)**

Supplementary Materials

Supplementary Table S1: Alternative Mortality Selection Corrections – Non-Parametric Cumulative Hazard vs. Logistic IPW

Age Group	Naive (Survey)	Cumulative Hazard (Stabilized)	Main IPW	% Δ Cumulative Hazard	% Δ IPW
25-34	0.00596	0.00521	0.01059	-12.60%	77.70%
35-44	0.02461	0.0237	0.03789	-3.70%	53.90%
45-54	0.05932	0.0576	0.08139	-2.90%	37.20%
55-64	0.10547	0.10352	0.12723	-1.80%	20.60%
65-74	0.16878	0.16562	0.1904	-1.90%	12.80%
75-84	0.20887	0.20534	0.23463	-1.70%	12.30%
85+	0.22836	0.2266	0.23955	-0.80%	4.90%

Supplementary Table S2: Sensitivity Analysis – IPW Corrections by Health Outcome Type

Age-group	Chronic Conditions Only		Acute Conditions Only	
	Survey Weights	IPW-Adjusted	Survey Weights	IPW-Adjusted
25-34	0.007*** (-0.001)	0.007*** (0.000)	.00 (0.000)	-0.001*** (0.000)
35-44	0.031*** (-0.001)	0.034*** (-0.001)	0.001** (0.000)	0.00 (0.000)
45-54	0.075*** (-0.002)	0.084*** (-0.001)	0.001* (0.000)	0.001** (0.000)
55-64	0.134*** (-0.003)	0.149*** (-0.001)	.001 (-0.001)	0.00 (0.000)
65-74	0.218*** (-0.006)	0.234*** (-0.002)	.001 (-0.001)	0.00 (0.000)
75-84	0.264*** (-0.010)	0.287*** (-0.005)	.00 (-0.001)	-0.001 (-0.001)
85+	0.300*** (-0.021)	0.300*** (-0.009)	.013 (-0.008)	0.003 (-0.002)
N	8,88,146	8,88,146	5,55,078	5,55,078