New Evidence on the Health-Status Gradient from Imperial China*

Karen Eggleston, Stanford and NBER
Wolfgang Keller, HKUST, U Colorado, CEPR, NBER
Carol H. Shiue, HKUST, U Colorado, CEPR, NBER

December 28, 2025

Abstract

We use Chinese family history data covering broad segments of the population to shed new light on the relationship between health and social status. In a sample centered on the early 1700s the typical lifespan of married men is 52, but varies widely, with a 90/10 range of 72 to 30 years. Status matters for health differences. High-status men have a 22 percent higher probability of living past 50 than low-status men, giving high-status men a lifespan advantage of seven years. For most of the sample period the health-status gradient is strongest at the top: men with lifespans ranging from the 20s to 60 all have comparable status, while only men with extraordinarily long lifespans achieve noticeably higher-status. This pattern begins to change in the 18th century as the gradient strengthens for men of lower status. Additional health-status findings are presented for the mens' wives and children. First, we show that high status is associated with even more survival advantage for the mens' wives than for the men themselves. Second, higher status translates into better health of children: the chance that the son of a high-status man dies before reaching 10 years is not even half of that of a low-status man's son. The findings provide new evidence on health-status inequality in a less-studied setting, and they also suggest that health-status inequality is usefully analyzed from a family and intergenerational perspective.

Keywords: Lifespan, Preston curve, health gradient, mortality, social status

*Ted Telford initiated the data collection underlying this work. Support from NIH is gratefully acknowledged (Shiue). Email: karene@stanford.edu, wolfgangk@ust.hk, shiue@ust.hk

1 Introduction

The negative correlation of income with premature mortality seems virtually ubiquitous -- it has been found wherever and whenever it has been sought (Deaton 2016), although pre-20th century evidence for non-western settings is limited. Precisely because its historical emergence is obscure, fully understanding the negative income-mortality correlation, aka Preston curve, has been a challenge.[1] This paper sheds new light on the negative income-mortality correlation using Chinese family histories spanning several centuries.

We analyze genealogical data from central China for births from 1298 to 1775, with the sample centering on the Qing dynasty (which started in 1644). Our primary health measure is adult lifespan. Social status can be distinguished based on more than 20 standardized descriptors, aggregated into four classes for a more parsimonious analysis. A key metrics of status are the educational attainment and official position of the male head of household, based on performance in the civil service examination system, with investments, occupation, household structure, and other indicators of wealth determining status as well. Consistent with population data for similar periods, the large majority of households never attained any status. Unlike historical studies using occupation to measure status in Europe, the Chinese civil service exam is a distinctive setting where male educational attainment directly conveys social status, and this status is not hereditary, but in principle and in practice open to a wide swath of the population.

Our data provides several advantages. Relative to most widely-cited historical cohort data (e.g. the Human Mortality Database, HMD), our data includes status and thus we can analyze the health-status gradient. Moreover, relative to earlier studies of China, which for data availability reasons tend to focus on the elite and special settings (e.g. Lee et al. 1994, Chen and Preston 2022), our sample includes more commoners and thus captures a fuller span of the status distribution in society.

Our analysis reveals a status gradient in adult lifespan of sizable magnitude. The positive health-status gradient during the sample period is about seven years difference in adult lifespan between the highest- and the lowest-status men. Equivalently, the chance of survival to 50 conditional on surviving to 15 (i.e., S(50|15)) was 72% for top-status graduates of the civil service exam (*jinshi*) and 58% for commoners. That range includes the S(50|15) of 63% for Swedish men born around 1750 from the Human Mortality Database, though their status distribution is unknown.

Generally, the health-status relationship is steeper for men of relatively high social status.

[1] Preston (1975).

In fact, men with lifespans ranging from 25 to 60 years typically do not vary much in terms of their status during most of the sample period. This pattern starts to change in the 18th century. Consistent with the hypothesis that the positive health-status gradient emerges from the top of the distribution, starting with birth cohorts post-1700, lifespan changes are closer to proportional in relation to status changes at all levels (i.e., closer to a linear health-status gradient).

We extend our health-status gradient analysis to two other family members: wives and children. Given family investments to find a suitable marriage partner, the relationship between wife lifespan and the status of her husband reflects not only the health-status gradient but also the extent of marriage matching.[2] We show, first, that the gradient between wife health and the status of her husband is not only positive but it appears to be stronger than the relationship between husband status and his own health. In particular, while highest-status husbands can expect to live seven years longer than the lowest-status husbands, their wives typically differ in lifespan by eleven years, in part reflecting higher maternal mortality among commoners. The average mortality rate between age 20 and 40 of the highest-status married women was half that of the lowest-status married women.

The gradient between father (husband) status and child health reflects intergenerational factors including family endowments and norms. Because mothers provide some of these, marital sorting is also reflected in the gradient. We find strong evidence for a gradient between household status and child health. Mortality of a son at any age–e.g., 5, 10, or 15 years–born to a top-status father is only half of what it is for a son of a father ranked just ten percentiles lower, and it is almost one third of the mortality of a son born to a low-status father. Our results on lifespan of wives and sons shed light on the role of the health-status gradient for the formation of kinship networks as well as intergenerational mobility.

While social status information is virtually always available, vital statistics are incompletely recorded in our data. In particular, information on year and month of death is missing for about one fifth of the men. It is common for samples during this period to be incomplete; for example, the widely-respected Human Mortality Database compiled by expert demographers does not include England and Wales until the 1840s due to incomplete vital statistics, and warns users of the earliest data – Sweden starting in the 1750s – to be cautious interpreting life tables before 1800 since they do not include unregistered individuals. Such incompleteness affects the gradient estimation because vital statistics are disproportionately missing for relatively low-status individuals subject to high mortality. In the present case, missing vital information can be estimated using auxiliary information or model lifetables

[2] A measure of a female's status based on her professional achievements or labor earnings is not available; this is typical for this historical period.

for historical periods, and our baseline includes such reconstituted observations. Results for more limited samples presented in the Appendix also show a health-status gradient, though its magnitude is smaller (and likely underestimated) due to the underrepresentation of lower-status men.[3]

Our analyses contribute important historical perspective for understanding of the strength of the health-status gradient. A large literature studies the relationship between socio-economic status (SES) and health (overviews include Cutler, Deaton, and Lleras-Muney 2006; Cutler, Lleras-Muney, and Vogl 2011; Deaton 2016). A frequent starting point for studying the gradient in different populations and time periods is the cross-country relationship between health and per capita income as described originally by Preston (1975), though within-country analysis, as in the present paper, has produced additional insights (Chetty et al. 2016). We present novel evidence about how adult lifespans differed by social status for a time and region with little previous evidence, but with great importance for understanding the well-being component of the 'great divergence' (Pomeranz 2001) between China and the west.

Unlike most previous studies, our cohorts spans several centuries, allowing us to examine the evolution and reproduction of the Preston Curve across multiple generations. We also contribute to the historical and economic literature by analyzing data with a relatively time-invariant, multiple-category metric of SES – so that we are documenting a gradient, not merely a gap between a binary stratification of rich versus poor, or elite versus commoner. As emphasized by Cutler, Vogl, and Lleras-Muney (2011) and others, the SES-health gradient reflects a bundle of factors in SES that potentially have distinct independent influences on health (e.g., short-term income fluctuations versus long-term income or wealth, or education which is not subject to short-term fluctuations). In our setting, these dimensions of status are deeply inter-connected through the social prestige of ranking on the national civil service examinations – higher educational attainment leading to stable employment, assortative matching with wealthier marriages for himself and his sons, and bureaucratic or social power over consequential community resources. The significant gradients in lifespan for married men are supplemented with evidence on differential wife lifespan and child mortality by household status. Moreover, Qing dynasty China represents an interesting period to study because of its dramatic population growth and stagnant or declining living standards; by contrast, many previous studies focus on later periods with better vital statistics, especially cohorts living through improving living standards and overall rising life expectancy in the 20th century.

[3]Data reconstitution generically leads to measurement error. For present purposes, it is plausible to assume that it is a price worth paying (see also Telford 1990).

We also contribute to a small but growing literature using genealogical data to study long-run questions of economic status or health. An early contribution is Fogel (1986), who compiled a sizable sample of genealogical records to better understand the evolution of mortality in North America. Recent work on Malthusian population responses in China using genealogical data includes Hu (2023), and data on lineages from this part of China have been used to address questions of the fertility quantity-quality trade-off, the evolution of social mobility as the return to human capital changes, as well as the long-run impact of big shocks on the power of elite (Shiue 2017, 2025, and Keller and Shiue 2024, respectively). Surveys of Chinese genealogies include Liu (1978) and Shiue (2016).

The results presented here are based on a small number of genealogies from a particular county in Qing China. Because broad administrative data for comparing our sample with population statistics is not yet available, it is impossible to know to what extent our results generalize to other parts of China. At the same time, census-like data that guarantees representativeness typically does not exist until the late 20th century; before, even studies with large samples are subject to unknown selection.[4] Because data availability might affect the estimation of historical health gradients, the representativeness of our sample is discussed in section 3. Relative to most studies for similar historical periods, here commoners constitute the majority of the sample, a key advantage for studying the health gradient. In addition, our analysis is unusual in that we provide a comparison of gradient findings for samples that vary in their reliance on reconstituted data (see the Appendix). This is important for gradient estimation because it allows the reader to pick their preferred scenario along the dimensions of sample selection and measurement error.

The paper is organized as follows. The next section describes the historical setting, followed by a description of data characteristics and representativeness. Section 4 presents our main empirical findings on the health-status gradient for men, while the following section extends them to married women and sons. The final sections discuss our results and briefly conclude. More information on the data and additional estimation results can be found in the Appendix.

[4] For example, recent work on 1880-1920 US includes no more than 1/3 of the target population, and 99% of the sample is white (Black et al. 2024). Reweighting the sample will only work if one has good reasons to believe that the 1% remaining non-whites in the sample are representative of all non-whites in the population. See also Bailey, Cole, Henderson, and Massey (2020) on sample composition in ex-post linked US census data.

2 Historical Setting

While this study covers (parts of) three imperial dynasties–the Yuan (1271-1368), Ming (1368-1644), and Qing (1644-1911)–the sample is centered on cohorts in the early 1700s, the relatively early Qing dynasty. Total population grew from about 200 million in the mid-17th century to 400 million by 1900 (Cao 2000). The effective power of government was organized around an autocratic central authority and lower administrative regions. The state taxed lightly in comparison with other states of the period. Moreover, the scope of markets for allocating scarce resources was limited by technology (e.g., transport technology, financial instruments) more than by government regulation, in part because effective enforcement would have required more state resources than were in fact allocated.

In this structure of governance by bureaucratic-scholar officials, education was a cornerstone of entry into higher-income classes and a key aspect of the social contract between the state and local management. Consistent with low central taxation, the capacity of the state was limited. The provision of public goods was instead delegated to local governments and the leaders of local clans.

Thus, local elites exercised paternal authority over their extended families, while the participation of local elites in the political decision-making body of the state was legitimized through the state-run civil service examination system. Throughout the Ming-Qing period, participation in the tournament-style civil service exams was the most direct path towards status and income. Passing the exam and obtaining a government office held substantial rewards. Social status derived from passing the civil service examinations was the "ultimate source of power" (Ho 1962). These returns were so attractive that non-scholar officials who obtained wealth also invested in their sons' education.

On the one hand, local leadership by elites who passed the civil service exam strengthened and ensured the legitimacy of the central state. On the other hand, local elites who supported the state through participating in the civil service earned high financial rewards. There thus arose, in effect, a partnership between the central state and local elites, to mutual advantage.

3 Data

3.1 Chinese Genealogies as a Source of Information

In this paper, we use one of the main sources of information produced by clans, the genealogy. Genealogies allow for a more expanded definition of status compared to many other sources of SES proxies, because they include not only status information mirroring the ranking of exam degrees and official titles used by the state, but in addition, genealogies record non-official types of wealth that clearly mattered.

Chinese genealogies are documents that record a family's history, providing information in the form of an annotated family tree where male individuals are listed in terms of an intergenerationally-linked structure along with the names and vital statistics of their spouses and children. Genealogies originated from beliefs surrounding ancestral ceremonies and Confucian teachings dating from at least the Tang Dynasty (618-907 A.D.). By the Ming and Qing, the practice had spread widely across the empire. As such, the Chinese genealogy is a well-known type of historical document that falls under the category of household documents (Faure and He 2020). Like other household documents—which also include contracts, prayers for religious rituals, religious manuals, account books, tax payment records, almanacs, and textbooks—genealogies preserved information that was important to families within the social order of late imperial China. Wang (2008) catalogs approximately 50,000 publicly available genealogies, the bulk of which were produced in the Ming and Qing Dynasties.

Because the documents are self-reported, issues of representativeness, selection bias, and accuracy need to be considered. Not every family had a genealogy; some historical genealogies have survived while others have not. Also, Chinese genealogies are organized patrilineally—the family tree follows the marriages of the males of the family—implying less information on daughters than for sons. Compilation also required that at least some family members were literate, in which case the set of families for which genealogies exist might be better off than the set of families for which they do not exist (wealth bias). The principle of Chinese genealogical compilation was such that all males in the patrilineal line were included for reasons of ritual.

Clan membership was self-professed, but it was important because it defined and clarified allegiances and responsibilities during times when conflict might create uncertainty regarding obligations. Family laws were not immutable, and they could adapt to new situations, but the recording of details about members in genealogical charts—biographies, obituaries, official degrees, appointments to office—reflected the lineage as a force for communal organization. Unlike official household registers that linked population counts to tax levies (and which resulted in a built-in disincentive for accurate reporting), members of clans were more invested in the accuracy in their genealogical documents.

Other economic functions of Chinese genealogies are related to the way China was governed during the sample period. Whether implicitly or explicitly, the state delegated much of the administration of local public goods to local clans. Tasks such as the management and upkeep of irrigation systems, poor relief, and grain storage for famines would oftentimes be passed down from the district magistrate to local clans. The amount that clans could contribute depended in turn on the resources at their disposal and the number of members in each clan. For example, lineage school fees were assessed on lineage members, with fees rising in an individual's income, providing reliable information on income differences between clan members (Chang 1955).

In summary, there is a considerable amount of information in family records that were privately maintained. Genealogies were common not only because of their centrality for ancestral rituals, but also because of their role documenting a written record that describes private forms of economic and political organization.

3.2 Coverage and Representativeness

The data are from genealogies of clans that resided in Tongcheng County of Anhui Province. In the late 18th century, Tongcheng county had about 1.3 million inhabitants (Beattie 1979). The county was relatively developed and densely settled, one of the many thriving regions located near the lower Yangzi River Basin. Tongcheng was about 150 miles away from Nanjing, the early Ming Dynasty capital; Beijing, the later Ming and Qing capital, was about 650 miles away. The region was mainly a rice-producing area where the wealthiest families were typically landowners. Over the Ming and Qing Dynasties, the region gained some fame for having produced a number of the highest officials of the empire.

Genealogies exist for numerous clans in Tongcheng, with one account relying on more than 60 genealogies (Beattie 1979), and another study utilizing 39 genealogies (Telford 1992). The sample we use is created from data collection initiated by Ted Telford in the 1980s. Because employing all known genealogies from Tongcheng would lead to the oversampling of elites, the present sample is based on a targeted approach, where our goal is to obtain a broadly representative sample in the raw data. Given the focus on lifespan and status, an important part of the approach is to target the fact that most of the population in the genealogy had no social status ("commoner") while top-status holders may have been around 2%.

To achieve a reasonably representative sample, we select the genealogies of seven clans, the Chen, Ma, Wang, Ye, Yin, Zhao, and Zhou. For the entire period covered in their genealogies, this amounts to a total of about 40,000 men, women, and children. Taken together, these genealogies yield a sample that is broadly representative for China as a whole. For an analysis of various aspects of selection and bias, see Shiue (2025).

Of the more than 11,000 couples recorded in the data, the present sample consists of the husbands born between 1298 and 1775. The majority of the sample lived during the late

Ming dynasty (1368 to 1644) and the early-to-mid Qing dynasty, when the principles of data recording became more standardized. We end the main sample with the 1775 birth cohort because the final year of compilation for some of the genealogies was in the 1860s; thus, using later birth cohorts would introduce censoring due to incompletely recorded lifespans.

In general, Chinese family genealogies vary in their completeness, and often some vital statistics are missing.[5] The extent to which the recorded data is subject to selection bias, as well as how it can be addressed, is discussed in Telford (1990). The baseline estimates in the text use vital statistics that are estimated using auxiliary information in the genealogies, as well as model life tables, a standard demographic tool. Model lifetables generally used for such reconstitution (e.g. Coale-Demeny, 1966 and 1989; United Nations 1982) are compilations of average age of death regardless of status that have been "designed primarily for use in developing countries or for estimating historic populations, [and] are limited to mortality patterns for a life span from age 20 to 75," with the Human Mortality Database (HMD) listed as the standard reference as the model lifetables were updated (https://www.un.org/development/desa/pd/data/model-life-tables). Application of such model lifetables to fill in missing vital statistics obviously reduces variation and omits the demographic impact of any region- or period-specific violence, famine, and epidemics. In the appendix, we discuss how data reconstitution affects our findings by presenting several sets of results with samples that exclude estimated data.

3.3 Measures of Status and Health

The genealogies record for each adult male his highest lifetime achievements in biographical sections that resemble obituaries. Our genealogies use more than 20 standardized descriptors of these achievements from which the social status of that man's household is derived. As is typical of data for the sample period, individual-level information is unavailable for other members of the household.

Participation in China's civil service examination, and based on that attaining official position, was key for a man's social status during the sample period (Ho 1962). One sign of the importance of the civil service exam for status is that high achievements in this dimension are virtually always recorded (Ho 1962). Other factors affecting social status in Qing China include occupation, education, making local investments, as well as other indicators of wealth and prestige. Table 1 shows summary statistics for the sample. These are all married men from the seven lineages that are born by the year 1775 with a completed lifespan of 16 or more years.

[5] An early survey of Chinese genealogies is Telford (1986).

Table 1: Summary Statistics

	N	Mean	Standard Deviation	Min	Max
Lifespan	4,656	52.39	15.50	16	91
Status (Percentile)	4,753	0.50	0.23	0.35	0.97
Total Number of Sons	4,753	2.12	1.61	0	10
Birth Year	4,754	1703.52	67.69	1298	1775
Year of Death	4,656	1755.66	69.82	1348	1856

Notes: Sample is all married men born on or before 1775 from seven Tongcheng clans.

The number of observations is N = 4,754. For virtually all of them we have information on maximal lifetime status. For robustness and to account for the size of each status class, we measure status in percentile rank following Shiue (2025) and others. Our analysis distinguishes four classes. The lowest status class covers 70 percent of the sample, with a class mid-point of 0.35, which is the minimum of our status variable. The highest social class has a mid-point percentile rank of 0.97, which means it covers men from the 94th to the 100th percentile. We see that the men have on average just over two sons, although that number varies from 0 to 10.

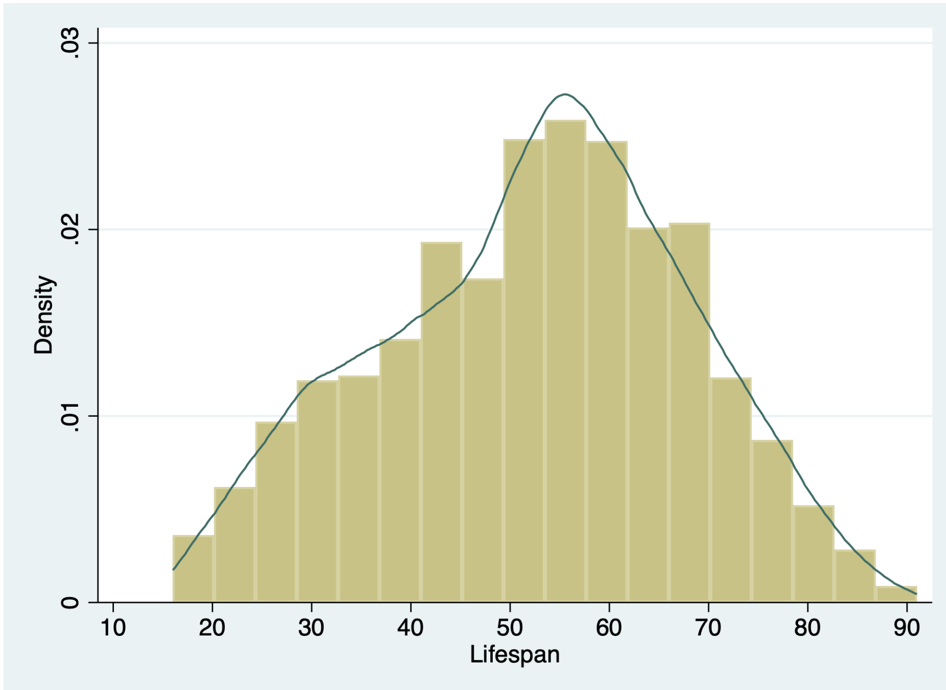
Figure 1 provides information on completed lifespan in the sample. With a 90/10 range in lifespan of 72 to 30 years, there is wide variation in lifespan. The interquartile range is 41 to 63.

The earliest birth year in the sample is 1298, although the majority of the sample men lived in the 16th and 17th century. We limit the sample to men born during or before the year 1775 to avoid problems associated with incomplete lifespan when the genealogies were updated for the last time. This constraint means that the most recent death in the sample took place in the mid-19th century (latest year of death is 1856). Notice that information on the year of death is less common than year of birth, and it determines for which of these men we know completed lifespan.

Table 2 gives information on the distribution of status.

Group 1 is the lowest status class, comprising of more than 70 percent of the sample. Those are individuals whose genealogical entry does not indicate any sign of elevated status ("commoners"). On the other end of the status distribution is Group 4 which includes graduates of the provincial and national civil service examination (*juren* and *jinshi*, respectively), as well as other high-status men. In between are men who had somewhat elevated status, for

Figure 1: Lifespan of Married Men in Seven Tongcheng Clans



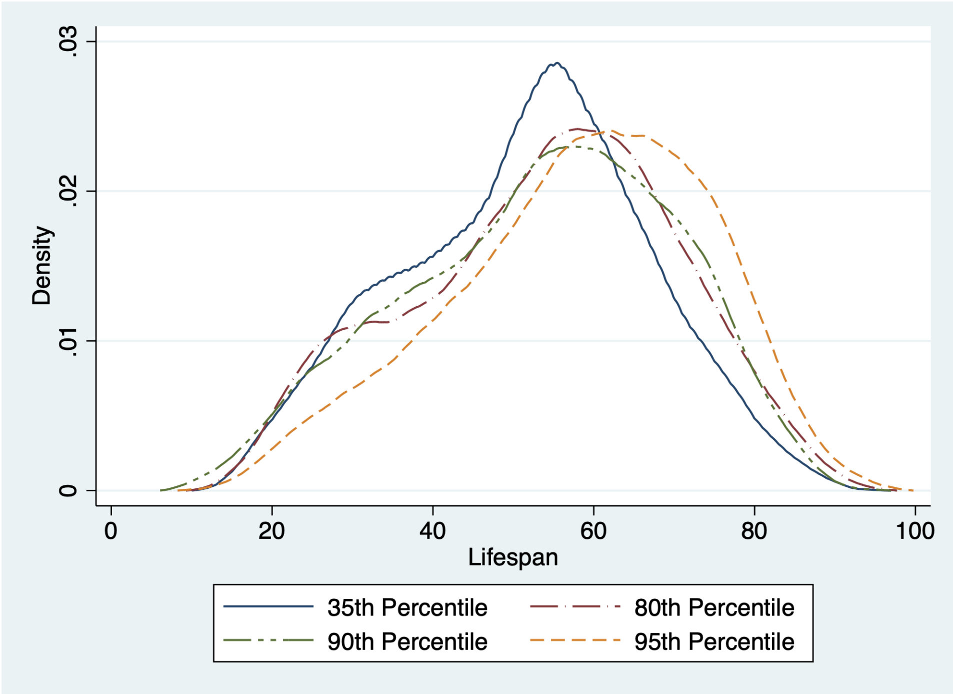
Notes: Completed lifespan of all married men from seven Tongcheng lineages with birth on or before 1775; n = 4,656.

Table 2: Status Groups and Lifespan

Status	Description	Percentile	N	Fraction of Sample	Lifespan (Mean)
1	No Status	0.354	3,368	0.709	51.41
2	Village head, multiple consecutive marriages, relative of minor official	0.782	693	0.146	53.79
3	Wealthy farmer, landowner, or merchant; scholar; official student	0.893	363	0.076	53.39
4	Graduates of the civil service examination at local, provincial (*juren*), and national (*jinshi*) level, high government officials	0.965	329	0.069	58.22
Total			4,753	1.000	52.39

Notes: Authors' calculation. Percentile of status is mid-point of status group; e.g. with 70.9% of men in the lowest status class, the mid-point is $(0 + 0.709)/2 = 0.354$. Social status coding informed by Chang (1955, 1962), Ho (1962) and Telford (1986, 1992). Number of observations (N) is for social status.

Figure 2: Lifespan Distribution by Status



Notes: Kernel estimates of completed lifespan by four social status levels given in Table 2.

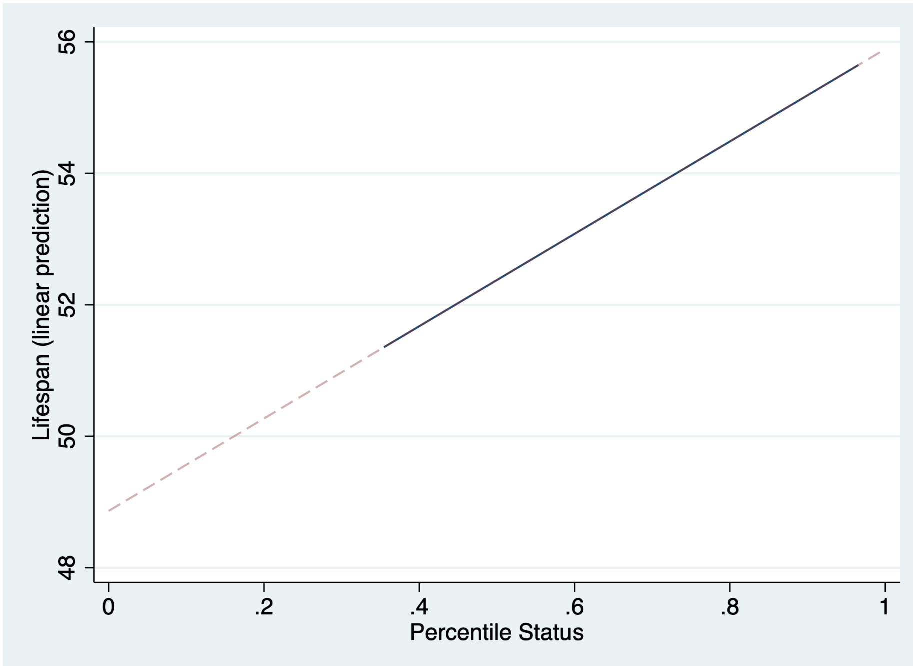
example because they were honored in village-level festivities (Group 2), as well as men who for example were wealthy enough to make significant donations and editors of genealogies (Group 3). This ranking of social status groups employed in this paper is informed by the work of Telford (1986), Chang (1955), Ho (1962), and others.

4 Results

We now turn to the relationship between health and status. Figure 2 provides densities of lifespan by status. We see that as status rises, the lifespan distribution tends to shift to the right. This is consistent with a positive health-status gradient. The difference is most clear between the lowest and highest status group. For men at the 95th percentile of the status distribution, there is a substantially fatter high tail in the lifespan distribution, while on the low lifespan side low-status men have a disproportionately high incidence of mortality in their 20s and 30s.

Figure 3 shows the Preston curve imposing a linear relationship. We see in Figure 3 that for this pre-1775 sample, there is a positive status-health gradient. The predicted change in lifespan for a man moving from lowest to highest status (0 to 100th percentile) is about seven years. A seven-year difference in adult lifespan is equivalent to half a century of longevity

Figure 3: Status-Health Relationship: Linear



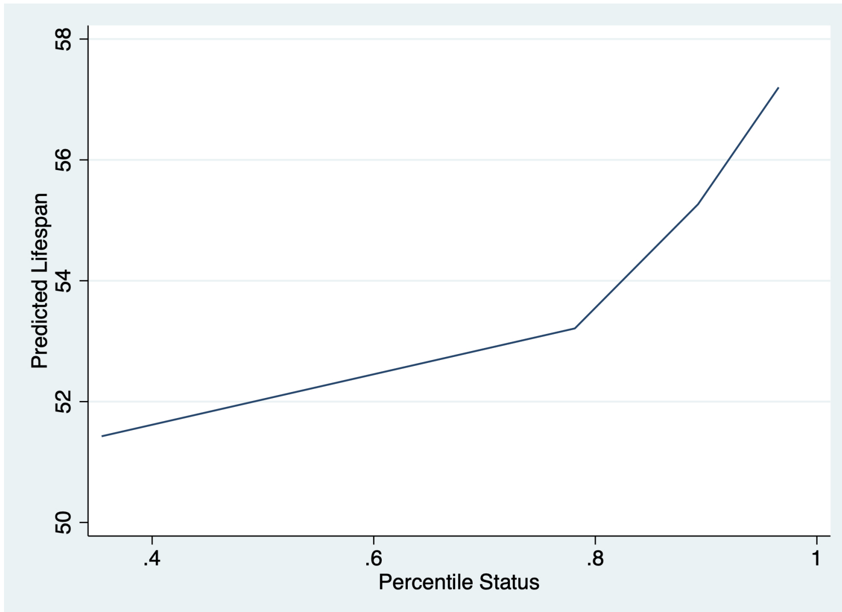
Notes: Linear relationship between percentile status and completed lifespan; n = 4,656; 95% percent confidence interval shown.

improvement in industrializing England and Wales, according to demographic data from the Human Mortality Database (HMD): the difference between the birth cohorts of 1841-44 and 1890-94 in England and Wales was 7.07 years lifespan (i.e., cohort life expectancy conditional on survival to age 15).

Allowing for non-linearities in the gradient, we see from Figure 4 that the status-health gradient in central China was steeper at higher status and health levels. These results parallel findings of a steeper gradient at the top end of the status distribution in the US during the early 21st century (Chetty et al. 2016).

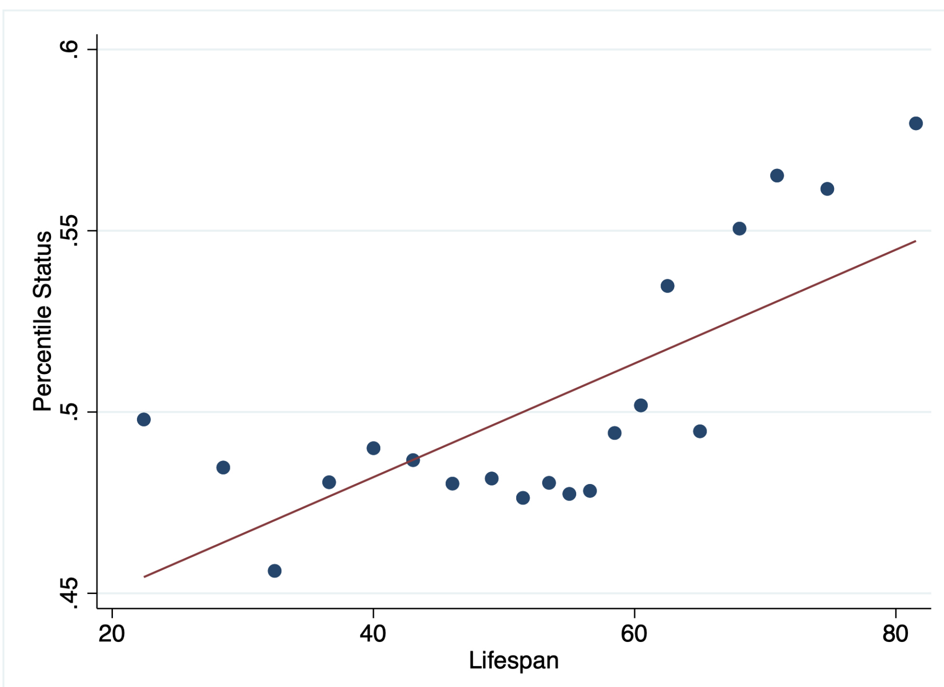
We cannot analyze potential non-linearities at the lowest end of the status distribution similarly, because status differentials among commoners cannot be further distinguished. An alternative way to analyze the health-status gradient is to ask whether men with different lifespans differ systematically in terms of their social status. In essence, this amounts to flipping the horizontal and vertical axes compared to the previous analysis. Results are presented in Figure 5, which shows a positive relationship between health on the horizontal and status on the vertical axis, consistent with a positive Preston curve. At the same time, we also see in Figure 5 that for men whose lifespan ranges from the low 20s to almost 60 years, percentile status hovers around 48 percent (just below the sample average of 50 percent). It is mostly men with lifespans above 60 years that have substantially higher levels of status.

Figure 4: Nonlinear Status-Health Relationship



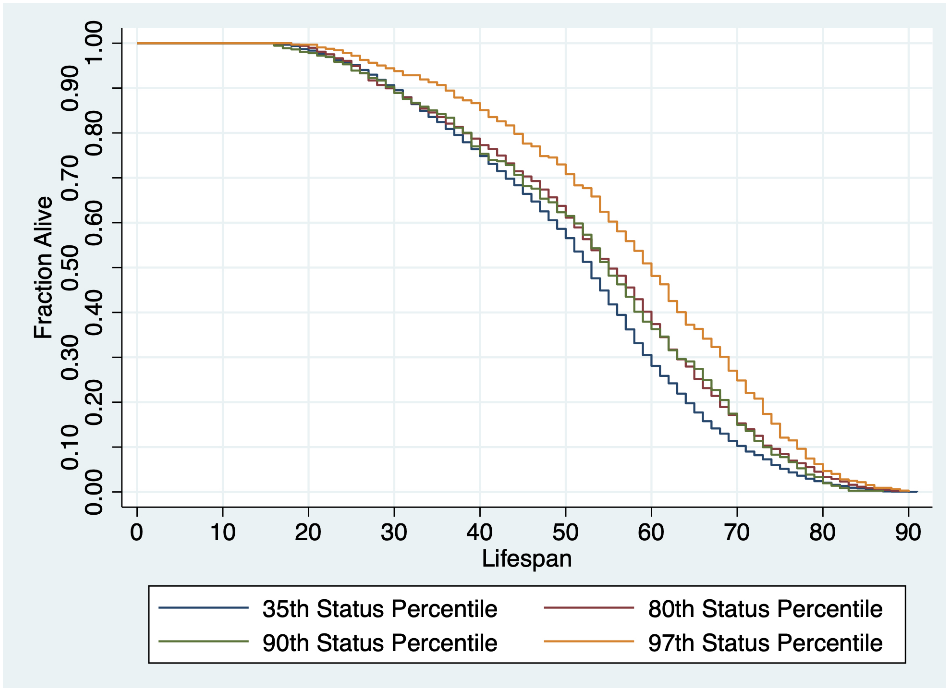
Notes: Fractional polynomial fit between status and lifespan.

Figure 5: Health-Status Relationship



Notes: Scatter plot showing average percentile status by lifespan bin; n = 4,656.

Figure 6: Survival by Status



Notes: Kaplan-Meier survival estimates by status; n = 4,656.

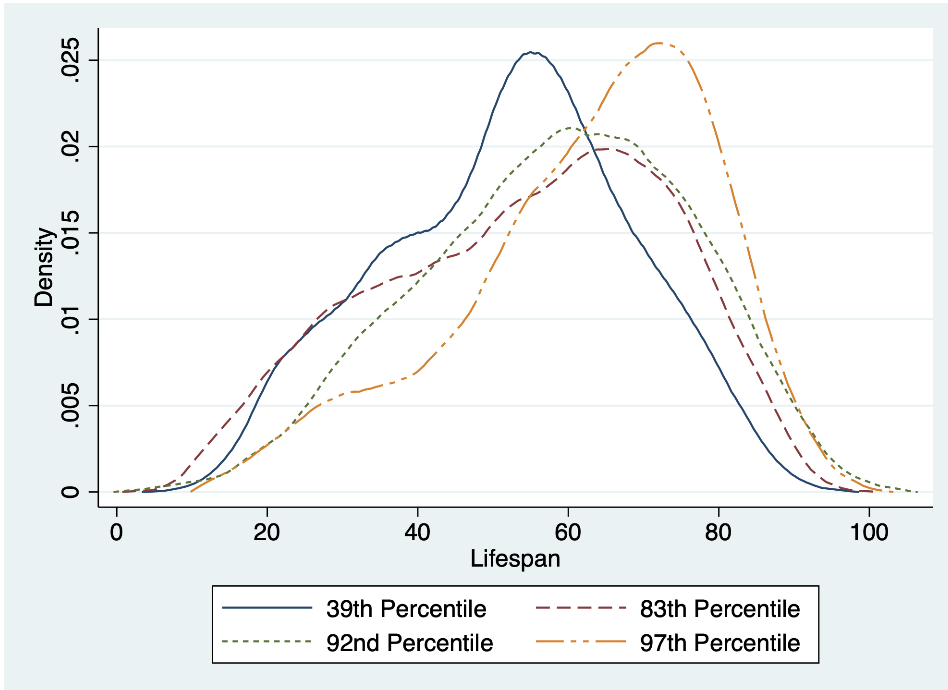
This finding confirms the earlier result that the Preston curve is steeper at higher levels of status.

A well-known tool for analyzing mortality differences is survival analysis. Figure 6 shows Kaplan-Meier survival estimates by status. There is a sizable lifespan gap between men at the 35th and 97th status percentile. For example, men of the highest status have a 72% chance of surviving to age 50, whereas only 58% of low-status men live longer than 50 years. Survival hazards for men at the 80th and 90th percentiles lie in between these extremes, as one would expect. This estimate is a plausible gradient that spans the 67.4% chance of survival to age 50 (conditional on surviving to age 15) for men born in England and Wales in 1841-1844, according to cohort lifetables compiled by the expert demographers of the Human Mortality Database (HMD).

5 Health-Status Gradient: Further Findings

This section extends our analysis of the positive health-status gradient in a number of dimensions.

Figure 7: Wife Lifespan by Husband Status



Notes: Distribution of wives lifespan by percentile status of their husbands; n = 3,706.

5.1 The Health-Status Gradient among Married Women

This section examines the health-status gradient further by deducing the female gradient from the lifespans of married women. We categorize women's household status based on her husband's maximum lifetime status. Given positive marital sorting, the husband's status and the wife's father's status are positively correlated, and high-status marriages partly reflect comparatively better health endowments. That is, one way of thinking about the lifespan of a married woman, given positive marital sorting, is that it gives additional information on the husband's health as a young adult, as well as that of the wife. The following results thus reflect both the degree of marital sorting as well as the underlying health-status gradient (if any).

Figure 7 shows the distribution of wife lifespan by the status of the husband.[6] Notice that there is strong evidence for a positive health-status gradient: for women with higher-status husbands, the lifespan distribution is shifted to the right. Mortality in the child-bearing years is clearly higher among lower-status women. In fact, at any given age between 20 and 40, high-status wives are twice as likely to survive as their commoner counterparts. This survival advantage reflects differential fertility and exposure to maternal mortality, as well

[6]This analysis focuses on women who are the single lifetime wives of their husbands.

as some protection against other health threats (e.g., from occupation or infectious disease) to which men were also exposed.Table 3 provides information on the size of each status group as well as the average lifespan of the married women in those household status groups. With a premium of almost 11 years for the highest compared to the lowest status group, the health-status gradient for women is steeper than for men (7 years difference). [7]

Table 3: Wife Lifespan and Husband Status

Status Group	Husband Status Percentile	N	Mean Wife Lifespan
1	0.393	2,870	51.95
2	0.836	389	54.14
3	0.917	241	58.51
4	0.974	205	62.72
Total		3,706	53.20

Notes: Percentile of status is the mid-point of the status group.

The strong female gradient is confirmed by Figure 8, showing the linear relationship between husband status and wife lifespan.

5.2 Child Mortality and the Health-Status Gradient

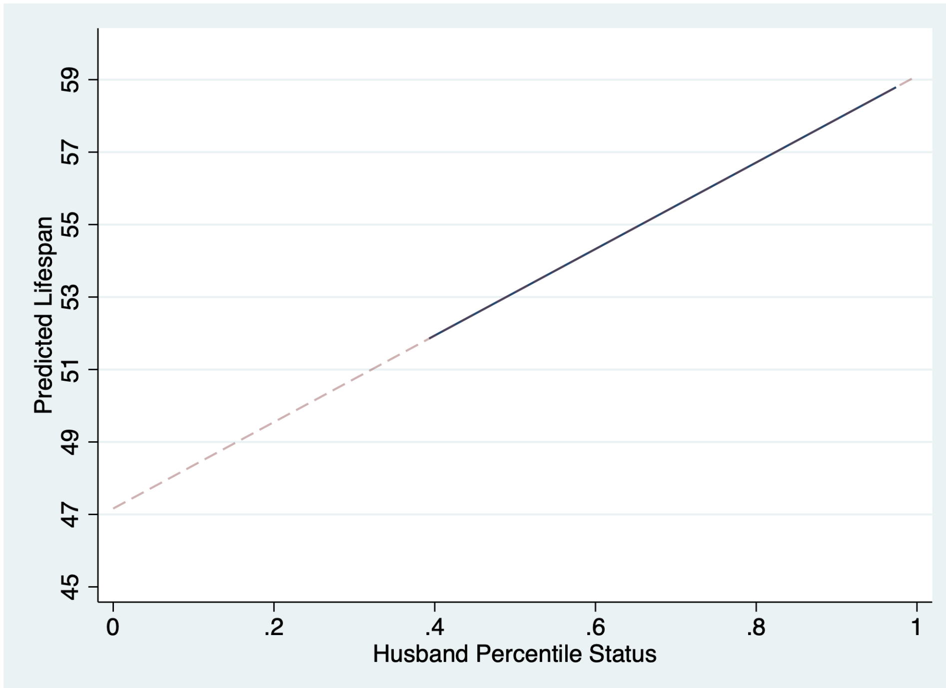
This section asks whether there is also a health-status gradient in terms of child mortality. For all married men born before 1775 from the seven clans, we examine the recorded lifespan of their male children. Figure 9 shows the distribution of their lifespans.

Several salient patterns emerge, related to our genealogical data. First, average lifespan of these sons is 41.5 years, which is more than 10 years less than the average lifespan of their fathers (see Table 1). This is because in the father generation, the analysis is conditional on survival to age 16 and marriage, whereas analysis of son lifespan includes individuals who will not survive to adulthood, as well as those men who stay single. In this analysis, for example, the fraction of sons who do not live past their fifth birthday (60 months) is 6.4 percent.

Second, even this formidable child mortality is likely an underestimate, given the moderate spike in deaths soon after birth compared to most historical records of premodern societies as well as for the Chinese elite. For example, infant and child mortality was about

[7]It should be kept in mind that the sample of men here is not the same as in the earlier analysis of male lifespan because that included men who had multiple female partners during their lifetime.

Figure 8: Linear Relationship between Husband Status and Wife Lifespan

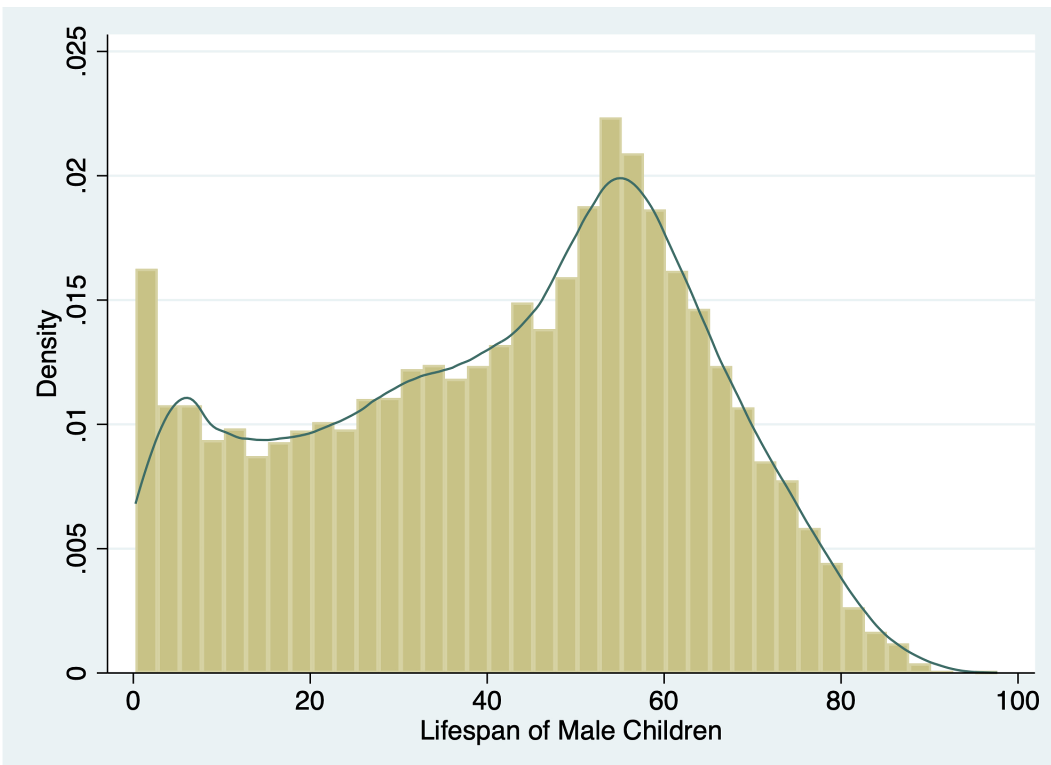


Notes: Distribution of married womens' lifespan in seven Tongcheng clans for women born in 1775 or before; n = 3,706.

10% among Qing nobility (Lee, Wang and Campbell 1994) and 8-10% for British elite boys (Hollingsworth 1957). The lower infant mortality rates (death within first 12 months of birth) in our data compared to estimates for this historical period are expected, given incomplete vital statistics especially for neonatal deaths as well as changing norms over the study period about when a son and his premature death should be recorded in the genealogy. (We know that recording of a child's birth is subject to considerable discretion, since baby girls were rarely recorded in the earliest period of our data, but are to some extent towards to the end.) Moreover, the incompleteness of infant mortality likely differs by status, given differences in fertility and in the mortality rates in youth. Accordingly, we will interpret the data conservatively as representing mortality after survival to age 1 (i.e., net of infant mortality). Any protection from infant mortality conveyed by high status (as hinted at in the data) will increase the gradient relative to what we document. Third, the lifespan difference between generations is consistent with a stagnant or downward trend in adult lifespan in late imperial China (Broadberry et al. 2018, Chen and Peng 2022), albeit conditional on (potentially improved) survival to age 15. In other words, the parental generation would have been subject to stronger selective pressure in childhood, and this selective survival contributed to the generational gap in adult lifespan.

In the following we examine the extent to which son mortality varies with the social status

Figure 9: Lifespan of Sons



Notes: Distribution of male childrens' lifespan of married men belonging to seven Tongcheng clans born 1775 or before; n = 9,406.

of their fathers. Table 4 presents results for mortality rates at different ages. We focus on survival to age 5, 10, or 15 in our exploration of the gradient and in comparing these Chinese patterns of child survival to those available for historical European populations.

Table 4: Mortality Rates of Sons

	N	One Year or less (12 months)	Five Years or less	Ten Years or less	Fifteen Years or less
34th Percentile Status	6,460	0.84	6.72	12.40	17.76
76th Percentile Status	1,447	1.11	7.32	12.65	16.10
88th Percentile Status	749	0.53	6.01	9.88	12.55
96th Percentile Status	745	0.13	2.95	4.56	6.58
All	9,401	0.80	6.46	11.62	16.20

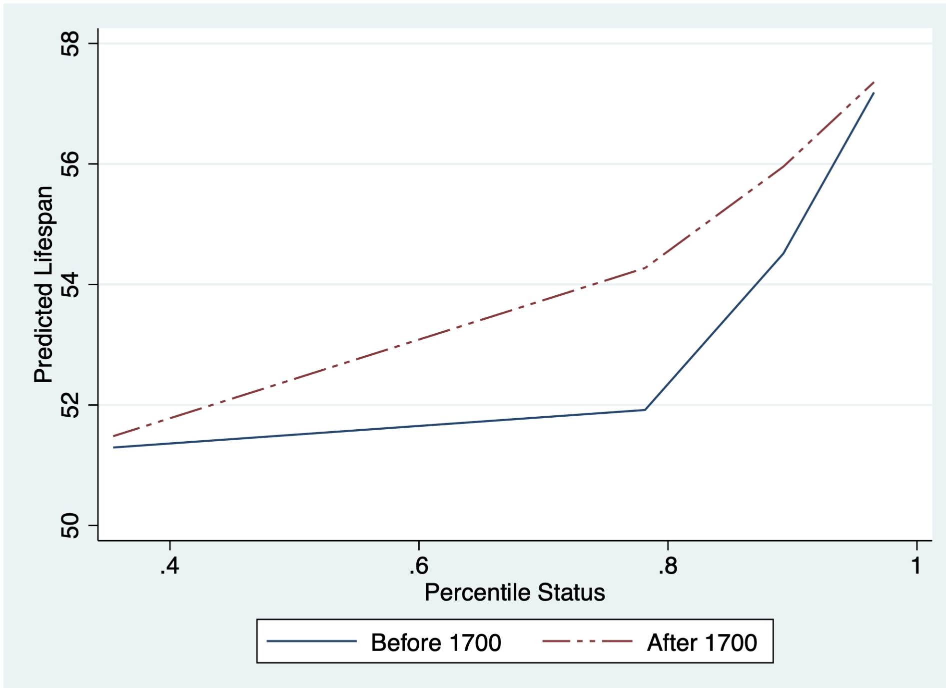
Notes: Mortality rates given in percent. number of sons for which father status is available: n = 9,401.

We see that consistent with a positive health-status gradient, mortality of these male children tends to be inversely related to the status of their fathers. For example, mortality by the age of 10 for the lowest status sons is 2.7 times the mortality rate of the highest status sons. Consistent with the steepening of the status-lifespan curve at high levels of status, we see that mortality rates are noticeably lower for sons from the highest-status families.

5.3 Status-Health Gradient Over Time

In this section we ask whether the status-health gradient has changed significantly over time. Two subperiods are distinguished; they are birth cohorts before the year 1700 and birth cohorts after 1700. Figure 10 shows the results, suggesting that the Preston curve has strengthened over time. In particular, while in the earlier subperiod it was mostly the top 10 percent of the status distribution who experienced longer lifespans, by the 18th century men at approximately the 80th percentile also showed signs of better health in the sense of a longer lifespan. Figure 10 is consistent with the diffusion in society of a positive status-health gradient from the top down, with health-protective behaviors or technologies (such as variolation for smallpox) invested in first by elites and later more broadly adopted or available. In ongoing research, we explore the role of intergenerational persistence in this diffusion of longer lifespans, and whether households experiencing downward mobility retain some protection from premature mortality relative to households that never attained status or literacy.

Figure 10: Status-Health Gradient over Time



Notes: Polynomial fit for birth cohorts before 1700 (n = 1,626) and birth cohorts 1700 and later (n = 3,029).

Robustness Accounts of the Ming-Qing period indicate that achieving high status became harder over time. To account for this we have examined the health-status gradient allowing for a time-varying status definition. Figure A.13 shows that this does not change our main findings.

6 Discussion

Analyzing unique data on lifespan and status across multiple generations for a sample centered on the 1700s, we show that there was a substantial health-status gradient in adult lifespan over centuries of Malthusian imperial China. Married men of the highest status lived on average 7 years longer than married commoners, and their wives' average lifespan exceeded that of commoners' wives by 11 years. These survival advantages represent 13% and 20% longer lives than the average lifespan for husbands and wives, respectively. Our findings contribute to the understanding of the positive correlation between health and status over a relatively long historical period where other evidence is sparse, especially for commoners (representing more than 60% of our data). We also document that for most of the sample period, the lifespan advantage of status was strongest at the top of the status distribution, with longer lifespans diffusing to lower-status households in the 18th century.

Few other studies document a pre-20th century health-status gradient in a non-Western setting, muchless during a period of rising and then static or falling living standards; the general increase in living standards and lifespan of the late 19th and early 20th centuries in Europe and the US are the subject of most previous studies (Preston 1975; Marmot et al. 1978; Cutler et al. 2011; Chetty et al. 2016; Black et al. 2024). Since our genealogical data primarily covers married men, the health and survival of other population subgroups would presumably demonstrate even larger disparities in health by status. Indeed, we present evidence of differences in child mortality by status that suggest differences in life expectancy at birth were probably larger that the 7-year gap in male lifespan between the highest- and lowest-status households.

While longer lives among high-status individuals may seem obvious and omnipresent, that has not always been the case. Long spans of history before the modern era appear to lack a significant survival advantage of the elite. For example, DeWitte et al. (2015) study London in 1700-1853, concluding that there were no evident gradients in adult health and survival, although lower SES households did suffer from higher infant and child mortality. Their assessment of little gradient in adult lifespan is consistent with numerous studies of disparate parts of industrializing Europe.

While evidence on cause of death during our study period is limited, we do know that elite lifestyles were not always conducive to longer lives: their food, though plentiful, did not necessarily reflect a balanced diet, and privileged behavior sometimes undermined health (e.g. overindulgence in alcohol). Many health-promoting behaviors are relatively inexpensive but time-consuming and may have been considered menial, such as fetching clean water and consistently boiling it before drinking. For example, although breastfeeding infants is protective of both infant and mother -- and poor households had few options but to carry on this tradition -- married women of ducal families of pre-1750 England employed wet nurses (Kendall et al. 2021): these elite English women had high fertility and closer child spacing, leading to lower immunity of infants and more susceptibility to early death from rampant infectious disease.[8]

To assess the magnitude of the lifespan gradient we find for imperial China, it is useful to

[8]More generally, the time price of health investments may lead to better health for those with lower opportunity costs of time, such as the poor. Ruhm (2000) showed that in recent US history, physical activity increases and diets improve during recessions, with less smoking and reduced obesity during periods of high unemployment with presumably lower opportunity cost of time. In an agricultural setting, Miller and Urdinola (2010) document higher child mortality in Colombia when world arabica coffee prices are high, a starkly procyclical pattern of child deaths linked to the opportunity cost of behaviors protective of child health.

compare to the earliest available and reliable cohort lifetables for Europe. We extract relevant survival percentiles from the earliest available cohort lifetables of European countries with good population-wide vital statistics, as compiled by the Human Mortality Database (HMD): for the 1700s (Sweden) and early 1800s (e.g. England & Wales). Cohort lifetables represent the appropriate counterpart to our data since they follow individuals born in a given year over a century until the cohort is extinct, in contrast to period lifetables (which are based on age-specific deaths in a given year). We use HMD lifetables over five-year periods to help smooth across historical fluctuations during this period.

This comparison shows that adult lifespan in this region of central China was comparable to several parts of Europe during the 1750s through 1840s, and that the adult gradient in survival by status was of the magnitude of several decades of European survival improvement. Specifically, a seven-year difference in adult lifespan is equivalent to half a century of longevity improvement in industrializing England and Wales, according to demographic data from the Human Mortality Database (HMD): the difference between England and Wales male birth cohorts of 1841-44 (the earliest available) and 1890-94 was 7.07 years lifespan (i.e., cohort life expectancy conditional on survival to age 15). Adult life expectancy was 58.04 for men born in England and Wales in 1841-1844, almost exactly the same as the average adult lifespan of the highest-status married men in our sample (58.22). The average lifespan of commoners (51) was slightly less than the 54.9 years on average that a man born in Sweden in the 1750s would live, conditional on surviving to age 15 (HMD, earliest cohort lifetable: Sweden 1751-1754). The chance of survival to age 50 conditional on surviving to age 15, S(50|15), was 63.5% for Swedish men born in the 1750s and 67.4% for the England and Wales male birth cohort of 1841-1844. We find that S(50|15) averaged 58% for low-status Chinese men and 72% for those of the highest status (conditional on survival to age 15 and marriage).

We also explore the genealogical data for evidence of differences in survival to adulthood by household status, given the high infant and child mortality in pre-modern populations. Studying the Qing nobility, Lee, Wang and Campbell (1994) estimate life expectancy at birth during this period was below 40. In the earliest available HMD data, Swedish life expectancy at birth was also below 40 throughout the 1700s, with at least 40% of babies dying before age 15. The chance of a baby boy surviving to age 10 conditional on surviving to age 1, S(10|1)), was 79.1% in Sweden 1751-54 and 82.2% in England and Wales 1841-44; these are equivalent to under-ten mortality rates of 20.9% and 17.8% respectively when infant mortality is removed (since we lack accurate data for infant mortality). These rates are a similar order of magnitude to the 11.6% under-ten mortality rate that we find. In England and Wales, male S(10|1) declined from 17.8% in 1841-44 to 8.8% in 1900-1904, comparable to the difference in under-ten mortality between a commoner (12%) and a high-status son

(5%).

These results for child mortality are suggestive that health may have played an important role intergenerationally, especially for higher-status households endeavoring to protect or improve their status. Controlling fertility to invest more in each son – a quantity-quality trade-off for male heirs – is less risky if child mortality can be kept low (and the mother would be more likely to survive the childbearing years to invest in the son's early education, enhancing cultural transmission). Other research has explored aspects of this mechanism. Shiue (2017) shows a negative correlation between fertility and education when the return to education was high; and Shiue and Keller (2024) explore the role of the quantity-quality trade-off in elite families' recovery from the devastation of the Ming-Qing transition. In further research, we study intergenerational persistence of lifespan and its relationship to status mobility in imperial China.

Our findings contribute novel results on the health-status gradient. Most previous studies of China investigate lifespans for a selective elite, with averages in the late 50s to 60s that are comparable to the highest-status men in our sample (Chen and Peng 2022, p.697; Lee and Li 2021). For example, Lee and Li (2021) utilize the recorded lifespans of 5000 extremely prominent Chinese elite over two millennia, from Confucius to high-profile officials, scholars, and generals as documented in the "Biographical Encyclopedia of Eminent Chinese in History"; their lifespans were relatively high (concentrated in the 60s and lower 70s) over the Ming and Qing dynasties. This lifespan range for Chinese elites of late imperial China is similar to that of high-status men with top-quartile lifespans in our sample. The comparative advantage of our data is the inclusion of a robust sample of non-elites, and thus the ability to document the health-status gradient for a broader swath of the population.

Studies documenting a status gradient in imperial China are even more limited. Analyzing genealogical data on the Qing dynasty nobility, Lee, Wang and Campbell (1994) find that infant and child mortality was similar in magnitude to that documented for European elites in the 18th century (e.g. 100 per 1,000 among Qing nobility, compared to Hollingsworth's data of 80 to 100 per 1,000 British elite boys, and other estimates for continental Europe; Hollingsworth 1957, Lee et al. 1994). Because reductions in child mortality may have led to stable or even increasing life expectancy at birth, our results are not inconsistent with assertions of Chen and Peng (2022) and others that overall human capital in China was remarkably high for low per capita income as well as quite stable despite exploding population. Indeed, Lee, Wang and Campbell (1994) estimate that improvement in child survival led to increasing life expectancy at birth of the Qing nobility during this period (from low 20s to about age 40), similar to or slightly earlier than among elite Europeans (Lee et al. 1994).

Potential mechanisms underpinning the differential survival by status and diffusion of health improvement from the elites to commoners remain unclear. Certainly elites in China like elsewhere had some incentives to help mitigate spread of infectious disease from widespread famine, and historical records indicate high-status households contributed charitable granaries as well as land to support educational expenses of poor young men, helping to diffuse literacy such that even some girls received some education (Beattie 1979, p.50). Clans also provided social protection to constituent households. Some additional candidate health production technologies include inoculation against smallpox and other traditional medical practices in China. Although the availability, cost, and uptake of variolation against smallpox in the populations we study is not well documented, this effective prevention strategy was widespread in many parts of China (and elsewhere) before spreading to Europe and North America. Evidence about smallpox inoculation during the 1700s indicates that differential uptake of variolation across the status distribution probably contributed to the gradient in infant and child survival, as well as to the significant improvement during the 1700s. Strong negative externalities of smallpox epidemics presumably spurred the diffusion of variolation across the status spectrum, as evident in the government-run variolation clinics in other localities during the period (Lee et al. 1994). There is less evidence about how other practices in Traditional Chinese Medicine (TCM) shaped the health patterns that we document. Most herbal medications are not extremely expensive and some TCM is well known to be quite effective against certain conditions. For example, a TCM-based therapy—artemisinin combination therapy—is now the globally recommended treatment for malaria. Further exploration of the mechanisms underlying the health-status gradient in imperial China would be valuable.

Limitations and extensions. Despite the many strengths of our data for studying lifespan and status, several limitations should be considered in interpreting our results and assessing their generalizability. The genealogies from Tongcheng represent the population in a small area of central China, and despite being more representative across status than many previously analyzed datasets, they need not be broadly representative of China's heterogeneous regions. We observe health and status along the male line, and generally only for those sons able to marry; although we can also parse the data to understand marital sorting, lifespans of married women, and some aspects of child mortality, the incomplete recording of live births and neonatal mortality, especially for girls, forces a focus on adult lifespans (i.e., conditional on survival to age 15). Nevertheless, the patterns we document are robust to limiting the sample to the most complete vital statistics and to accounting for changes in the cost of achieving status. These patterns suggest that status was somewhat protective of survival for children and mothers as well as for married men, and that the overall health-status gradient in life expectancy was larger than for adult lifespan. In ongoing work we are documenting intergenerational persistence in lifespan and its relationship with status mobility.

7 Conclusion

Studying imperial China with a sample concentrated in the 17th and 18th centuries, we document the evolution of the health-status gradient across multiple generations. Based on rich genealogical data that includes clear markers of status defined by China's national civil service examination system as well as occupation and wealth, we find that compared to married men without any evidence of wealth or status, those who achieved the highest rank lived on average seven years longer. This gradient in adult lifespan by status was of comparable magnitude to several decades of survival improvement in industrializing England. Some suggestive evidence also supports a gradient in female and child mortality over this period, indicating that life expectancy at birth may have exhibited an even stronger status gradient. These findings provide new evidence on the health-status gradient for a less-studied, non-western setting, as well as highlight the usefulness of an intergenerational household perspective for understanding inequality in health, socioeconomic status, and their inter-linkage.

References

[1] Bailey, Martha, Connor Cole, Morgan Henderson, and Catherine Massey. "How Well Do Automated Linking Methods Perform in Historical Samples? Evidence from New Ground Truth." Journal of Economic Literature 58, no. 4 (2020): 997-1044.

[2] Baten, Joerg, Debin Ma, Stephen Morgan, and Qing Wang. "Evolution of living standards and human capital in China in the 18–20th centuries: Evidences from real wages, age-heaping, and anthropometrics." Explorations in Economic History 47, no. 3 (2010): 347-359. References

[3] Beattie, Hilary. Land and Lineage in China: A Study of T'ung-ch'eng County, Anhwei, in the Ming and Ch'ing Dynasties. New York: Cambridge University Press, 1979.

[4] Black, Sandra E., Neil Duzett, Adriana Lleras-Muney, Nolan G. Pope, and Joseph Price (2024), "Intergenerational Transmission of Lifespan in the US", NBER wp 31034.

[5] Broadberry, Stephen, Hanhui Guan, and David Daokui Li. "China, Europe, and the great divergence: a study in historical national accounting, 980–1850." The Journal of Economic History 78, no. 4 (2018): 955-1000.

[6] Chang, Chung-li. The Chinese Gentry. Seattle: University of Washington Press, 1955.

[7] Chang, Chung-li. The Income of the Chinese Gentry. Seattle: University of Washington Press, 1962.

[8] Chen, Yuyu, Suresh Naidu, Tinghua Yu, and Noam Yuchtman. "Intergenerational mobility and institutional change in 20th century China." Explorations in Economic History 58 (2015): 44-73.

[9] Chen, Zhiwu, and Kaixiang Peng, "Production, Consumption, and Living Standards," Chapter 18, The Cambridge Economic History of China, Volume 1: To 1800 Edited by Debin Ma and Richard von Glahn. Cambridge: Cambridge University Press, 2022, pp. 676-709.

[10] Chetty, Raj, Michael Stepner, Sarah Abraham, Shelby Lin, Benjamin Scuderi, Nicholas Turner, Augustin Bergeron, and David Cutler. "The association between income and life expectancy in the United States, 2001-2014." JAMA 315, no. 16 (2016): 1750-1766.

[11] Cutler, David, Angus Deaton, and Adriana Lleras-Muney. "The determinants of mortality." Journal of economic perspectives 20, no. 3 (2006): 97-120.

[12] Cutler, David, Adriana Lleras-Muney, and Tom Vogl, "Socioeconomic status and health: dimensions and mechanisms." Chapter 7 in Oxford Handbook of Health Economics 2011, pp124-163.

[13] Dahl, Molly and Thomas DeLeire. "The association between children's earnings and father's lifetime earnings: estimates using administrative data." IRP Discussion Paper 1324-08, 2008.

[14] Deaton, Angus (2016), "On Death and Money: History, Facts, and Explanations", Journal of American Medical Association, April 26, 2016.

[15] Faure, David, and Xi He. "Family and Lineage in Late Imperial China", Oxford Research Encyclopedias, Asian History, 2020. https://doi.org/10.1093/acrefore/9780190277727.013.394 (Accessed November 1, 2023).

[16] Hollingsworth, T. (1957), "A demographic study of the British ducal families", Population Studies 11(1), 4-26.

[17] Ho, Ping-ti. The Ladder of Success in Imperial China: Aspects of Social Mobility (1368-1911). New York: Columbia University Press, 1962.

[18] HMD. Human Mortality Database. Max Planck Institute for Demographic Research (Germany), University of California, Berkeley (USA), and French Institute for Demographic Studies (France). Available at www.mortality.org (data downloaded on 7 June 2025).

[19] Keller, Wolfgang, and Carol H. Shiue, Intergenerational Mobility of Daughters and Marital Sorting: New Evidence from Imperial China, NBER Working Paper No. 31695, September 2023.

[20] Kendall, Ellen J., Alex T. Brown, Tim Doran, Rebecca Gowland, and Richard Cookson. "Health inequality in Britain before 1750." SSM-Population Health 16 (2021): 100957.

[21] Lavely, William. Review of Fate and Fortune in Rural China: Social Organization and Population Behavior in Liaoning, 1774-1873, by James Z. Lee and Cameron D. Campbell, Cambridge University Press, 1997, in American Journal of Sociology Vol. 104, no. 1 (1998): 256-258.

[22] Lee, James, Wang Feng, and Cameron Campbell. "Infant and child mortality among the Qing nobility: Implications for two types of positive check." Population Studies 48, no. 3 (1994): 395-411.

[23] Lee, James Z., and Wang Feng. One quarter of humanity: Malthusian mythology and Chinese realities, 1700–2000. Harvard University Press, 2001.

[24] Liu Ts'ui-jung. "Chinese genealogies as a source for the study of historical demography", In Studies and essays in commemoration of the golden jubilee of Academia Sinica. Taipei: Academia Sinica, 1978.

[25] Maddison, Angus, Chinese Economic Performance in the Long Run, 960–2030 AD, 2nd ed., revised and updated (Paris, OECD Publishing, 2007).

[26] Marmot, Michael G., Geoffrey Rose, Martin Shipley, and Peter J. Hamilton. "Employment grade and coronary heart disease in British civil servants." Journal of Epidemiology & Community Health 32, no. 4 (1978): 244-249.

[27] Maystadt, Jean-Francois, and Giuseppe Migali (2021), "The transmission of health across 7 generations in China, 1789-1906", Journal of Health Economics.

[28] Miller, Grant. "Women's suffrage, political responsiveness, and child survival in American history." Quarterly Journal of Economics 123, no. 3 (2008): 1287-1327.

[29] Miller, Grant, and B. Piedad Urdinola. "Cyclicality, mortality, and the value of time: The case of coffee price fluctuations and child survival in Colombia." Journal of Political Economy 118, no. 1 (2010): 113-155.

[30] Perkins, Dwight H. Agricultural Development in China, 1368-1968. Chicago, Aldine publishing, 1969.

[31] Pomeranz, Kenneth. The Great Divergence: China, Europe, and the Making of the Modern World Economy. Princeton University Press, 2000.

[32] Preston, Samuel H. "The changing relation between mortality and level of economic development." Population Studies 29, no. 2 (1975): 231-248.

[33] Ruhm, Christopher J. "Are recessions good for your health?" Quarterly Journal of Economics 115, no. 2 (2000): 617-650.

[34] Shiue, Carol H. "Local granaries and central government disaster relief: moral hazard and intergovernmental finance in eighteenth-and nineteenth-century China." The Journal of Economic History 64, no. 1 (2004): 100-124.

[35] Shiue, Carol H. "The political economy of famine relief in China, 1740-1820." Journal of Interdisciplinary History 36, no. 1 (2005): 33-55.

[36] Shiue, Carol H. "A culture of kinship: Chinese genealogies as a source for research in demographic economics." Journal of Demographic Economics 82, no. 4 (2016): 459-482.

[37] Shiue, Carol H. "Human capital and fertility in Chinese clans before modern growth." Journal of Economic Growth 22 (2017): 351-396.

[38] Shiue, Carol H. (2025), "Social Mobility in the Long Run. An Analysis of Tongcheng, China, 1300 to 1900", Journal of Economic History.

[39] Shiue, Carol H., and Wolfgang Keller. "Markets in China and Europe on the Eve of the Industrial Revolution." American Economic Review 97, no. 4 (2007): 1189-1216.

[40] Shiue, Carol H., and Wolfgang Keller (2024), "Elite Capital Strategies for Big Shocks: The Case of the Fall of the Ming", NBER wp.

[41] Telford, T. A. (1986). A survey of social demographic data in Chinese genealogies. Late Imperial China, 7, 118–48.

[42] Telford, T. A. (1990). Patching the holes in Chinese genealogies. Late Imperial China, 11(2), 116–35.

[43] Telford, Ted A. "Covariates of Men's Age at First Marriage: The Historical Demography of Chinese Lineages", Population Studies, 46, no. 1 (1992): 19-35.

[44] Telford, T. A. (1995). Fertility and population growth in the lineages of Tongcheng county. In S. Harrell (Ed.), Chinese historical microdemography (Berkeley: University of California Press), pp. 1520–1661.

[45] Wang, Feng, Cameron D. Campbell, and James Lee. "Agency, hierarchies, and reproduction in northeastern China, 1789 to 1840." In Prudence and Pressure: Reproduction and Human Agency in Europe and Asia, 1700-1900, p. 287. 2010.

[46] Weir, David R. "Family income, mortality, and fertility on the eve of the demographic transition: A case study of Rosny-Sous-Bois." The Journal of Economic History 55, no. 1 (1995): 1-26.

Online Appendix

A Data

A.1 Genealogical Data

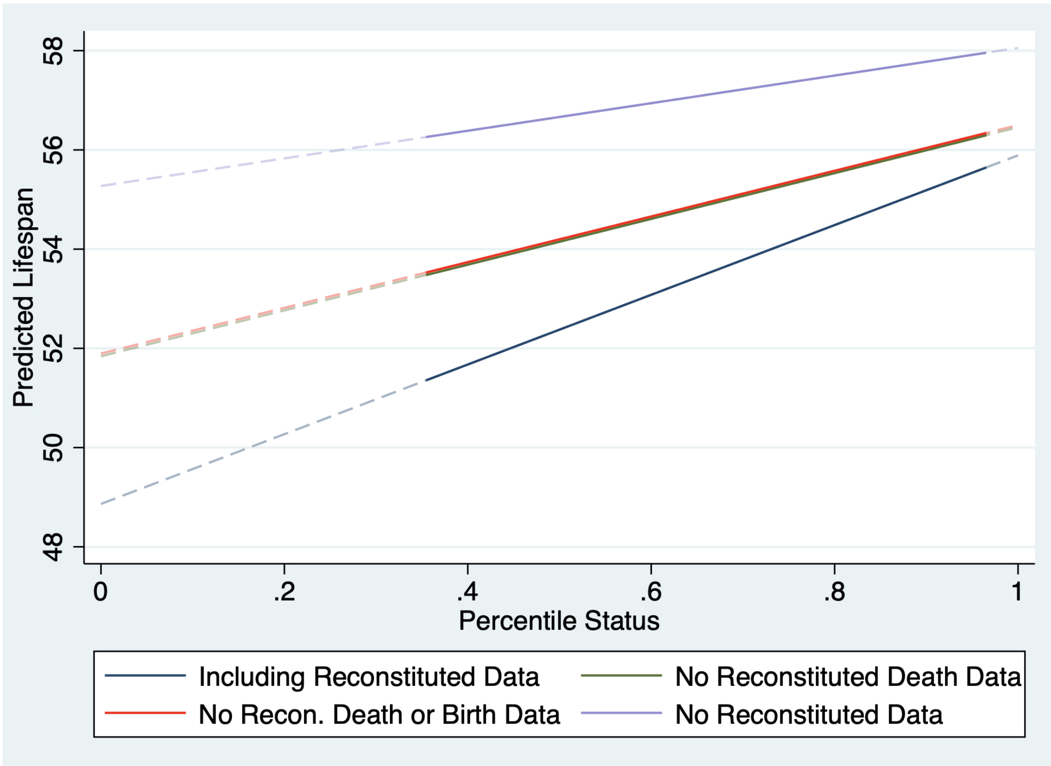
The sample comes from Tongcheng county in Anhui province. The following discusses the most salient features, with more details given in Shiue (2025).

B Empirical Results for Alternative Samples

B.1 The Role of Data Reconstitution: Introduction

To avoid selection bias, our baseline sample includes in part reconstituted vital statistics. Fortunately, the influence of data reconstitution on the results can be directly assessed because we know which data is reconstituted and which is not. The following shows evidence on the linear health-status gradient for a range of samples for comparison, see Figure A.1.

Figure A.1: Status-Health Relationship: Baseline vs Alternative Smaller Samples



Notes: Linear relationship between percentile status and completed lifespan; Including Reconstituted Data (Baseline): n = 4,656; No Reconstituted Death Data: n = 3,540; No Reconstituted Death or Birth Data: n = 3,496; No Reconstituted Data: n = 2,888.

Figure A.1 shows the baseline result from the text on the bottom (Including Reconstituted Data); the predicted lifespan difference between highest and lowest social status (100th versus 0 percentile) is about seven years. If observations with estimated death date are dropped (n = 1,069), we obtain the line located in the middle of Figure A.1. The predicted lifespan difference is now just under five years.[9] If in addition we drop observations for which the birth date but not the death date is estimated, this yields very similar results (largely because there are only n = 44 observations for which the birth but not also the death date is estimated).

Finally, if we drop in addition also the n = 705 observations for which both birth and death data is estimated, we obtain the linear health-status gradient shown at the top of Figure A.1. This is the most conservative sample definition designed to reduce measurement error. The predicted lifespan difference between highest and lowest-status men is now less than three years. This higher, flatter gradient compared to the baseline gradient in Figure A.1 reflects the fact that in the unreconstituted sample, high status men are overrepresented (see Telford 1990).

[9]Here we also drop 47 observations for which vitals are adjusted to correct for typos.

The following section presents additional results comparing the most conservative with the baseline sample.

B.2 Additional Results for Unreconstituted Sample

Table A.1 shows summary statistics for the sample using unreconstituted data.

Table A.1: Summary Statistics

	N	Mean	Standard Deviation	Min	Max
Lifespan	2,817	56.77	15.17	16	91
Status (Percentile)	2,888	0.50	0.25	0.32	0.95
Total Number of Sons	2,888	2.39	1.63	0	10
Birth Year	2,889	1712.65	59.95	1330	1775
Year of Death	2,817	1768.92	60.78	1380	1856

Notes:

Table A.2 gives information on the distribution of status in the limited sample.

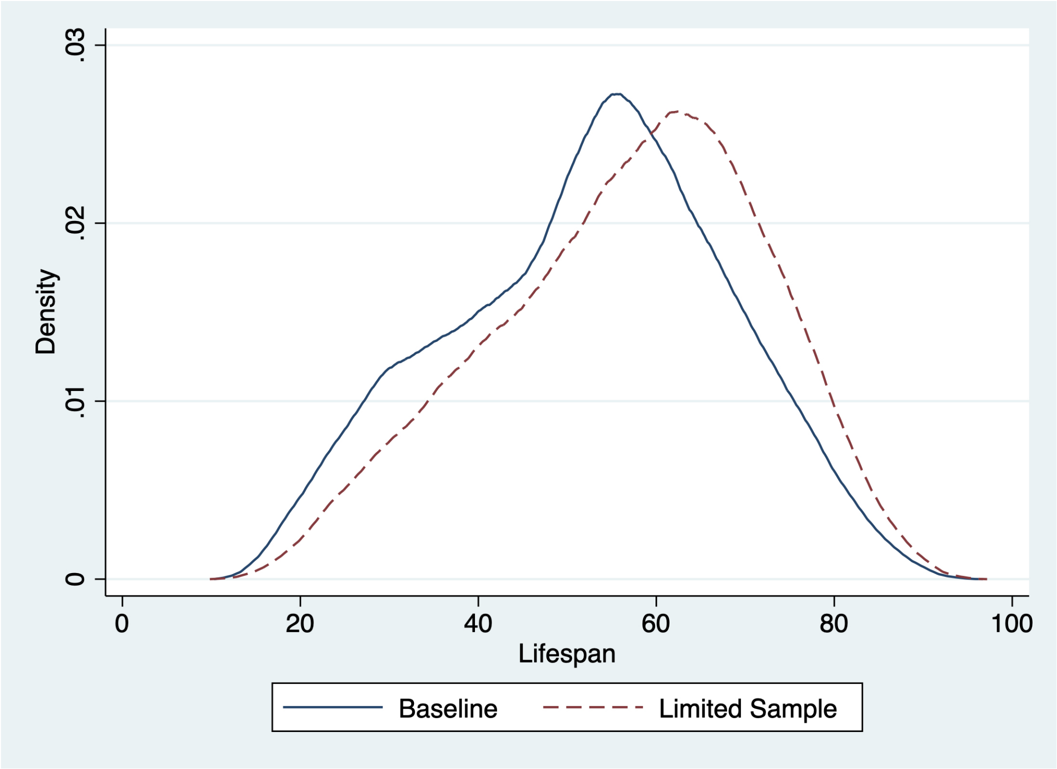
Table A.2: Distribution of Status

Status Group	Percentile (mid-point)	N	Fraction of Sample
1	0.321	1,854	0.642
2	0.725	476	0.165
3	0.851	257	0.089
4	0.948	301	0.104
Total		2,888	1.000

Notes:

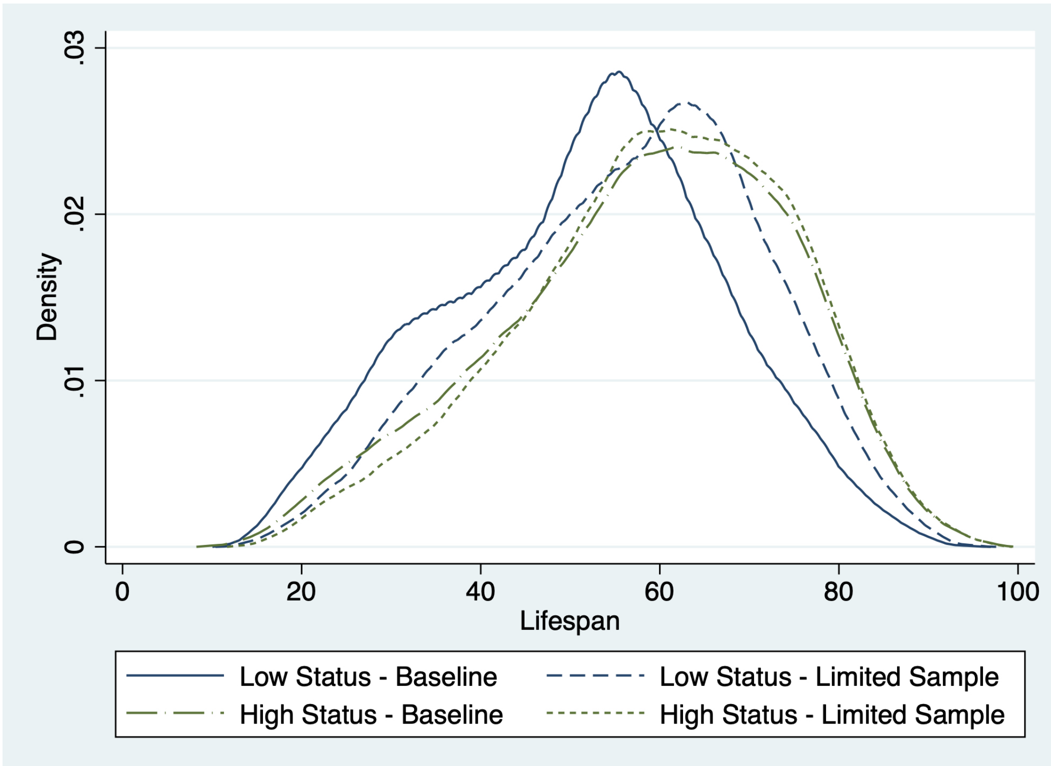
Notice that the composition of status in this sample is shifted to higher-status men, compared to the baseline sample. For example, the share of lowest-class men is 64 percent, in contrast to 71 percent in the baseline sample. This is because poorer (lower status) men are more likely to have vital statistics missing in the genealogies, see Telford (1992). In addition, data reconstitution also raises the sample size by about 40 percent (n = 4,753 versus n = 2,888). Because in this unreconstituted sample high status men are overrepresented, Table A.1 and Figure A.2 show that the distribution of male lifespan is shifted towards longer lives, yielding a mean difference of about 3.5 years. Importantly, this shift is not symmetric across all lifespans. In particular, the limited sample does not show the propensity of men to die around 30 years of age that we see in the baseline sample. The main reason for this difference is that the limited sample omits men who die at a relatively early age during the Ming-Qing transition.

Figure A.2: Lifespan Distribution: Baseline versus Limited



Notes: Completed lifespan of married men from seven Tongcheng lineages with birth on or before 1775. baseline n = 4.656, limited sample n = 2,888.

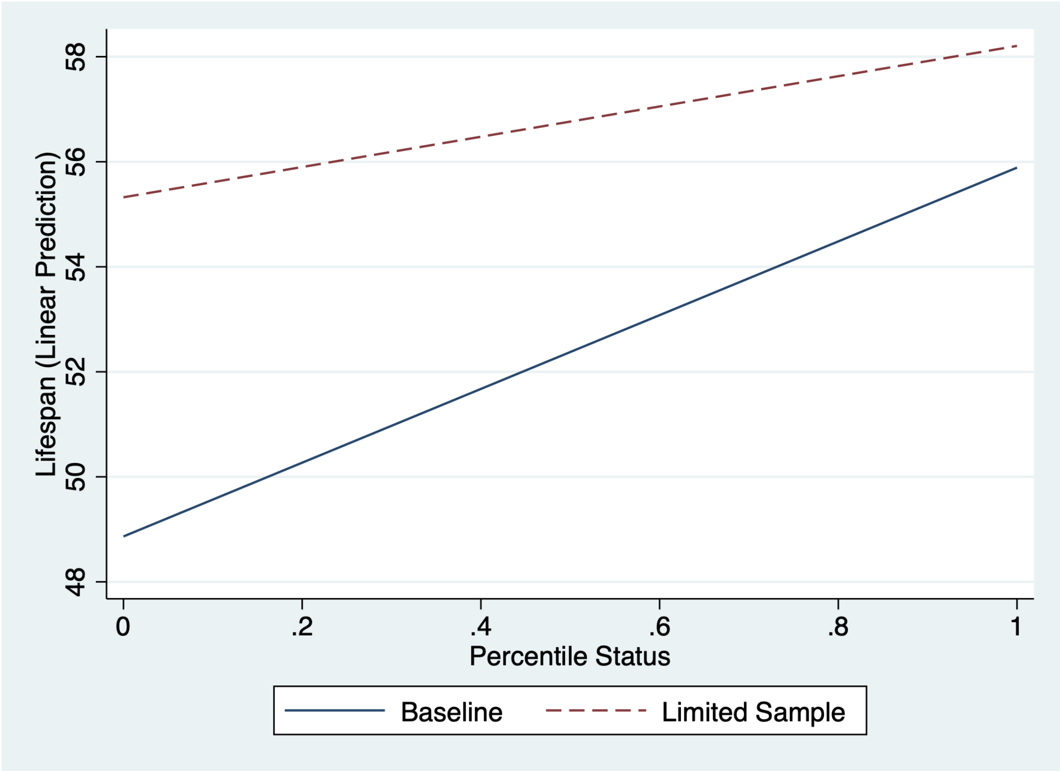
Figure A.3: Lifespan Distribution by Status: Limited vs Baseline



Notes: Kernel estimates of completed lifespan by highest and lowest status levels for baseline and limited sample.

Figure A.3 provides lifespan densities for the highest and lowest status groups, comparing the baseline and limited samples. We see that there is a greater difference in lifespan densities between baseline and limited sample for low status than for the high status men. This confirms that vital statistics that are unavailable in the limited sample tend to be for lower status men. Figure A.4 shows the linear Preston curve separately for baseline versus limited sample.

Figure A.4: Status-Health Relationship: Baseline vs Limited Sample

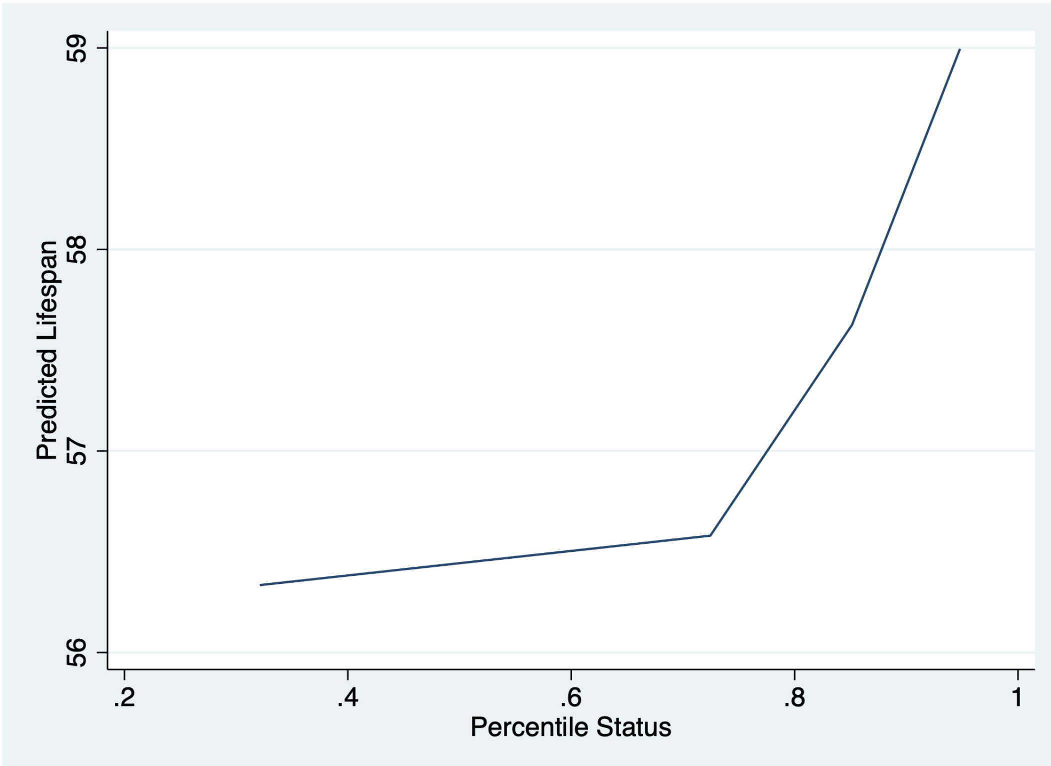


Notes: Linear relationship between percentile status and completed lifespan; n = 4,656 in baseline and n = 2,888 in limited sample.

We see from Figure A.4 that there are two main differences between the linear Preston curves for the baseline and limited samples. First, the Preston curve for the smaller sample is less steep than for the baseline sample; the predicted change in lifespan for a 100 percent change in status is only 2.5 years, instead of 7 years for the baseline sample. Second, the location of the Preston curve for the limited sample is at a higher lifespan level. Both of these features are related to the fact that vitals for relatively low status men are disproportionately missing. At the same time, results in Figure A.4 indicate that a positive gradient exists irrespective of sampling questions.

Figure A.5 indicates that the gradient is particularly steep at high status levels not only for the baseline but also for the limited sample.

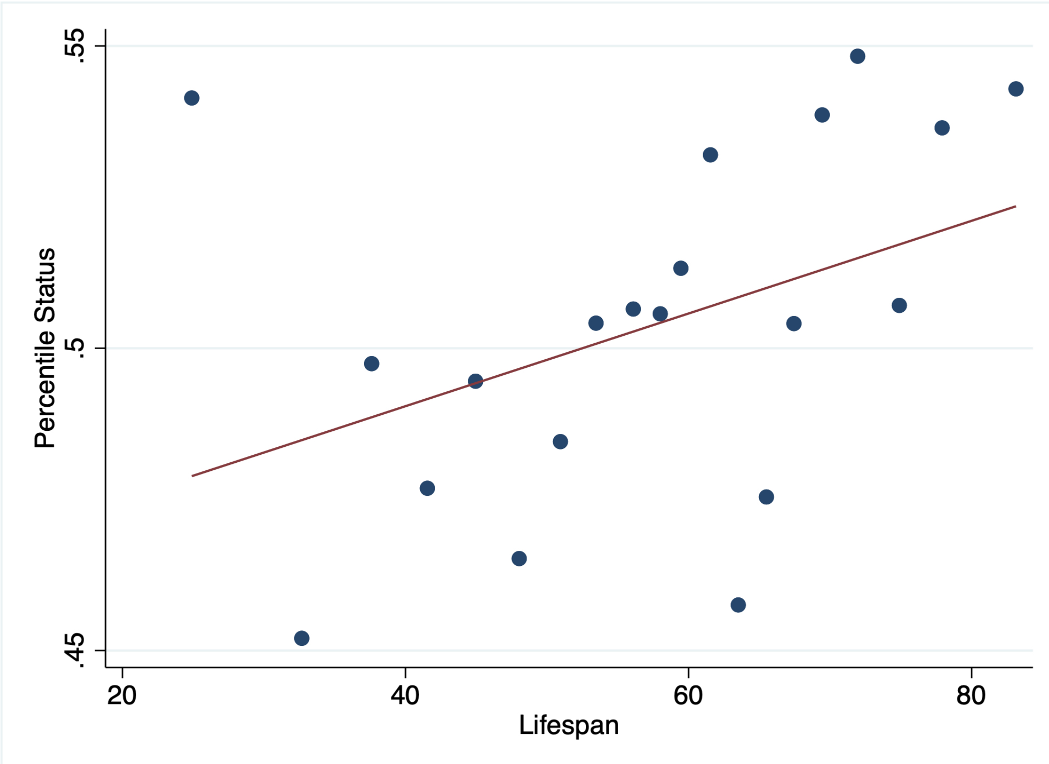
Figure A.5: Nonlinear Status-Health Relationship: Limited Sample



Notes: Fractional polynomial fit between status and lifespan; n = 2,888.

Figure A.6 presents the binscatter plot between lifespan and status for our limited sample.

Figure A.6: Health-Status Relationship in Limited Sample

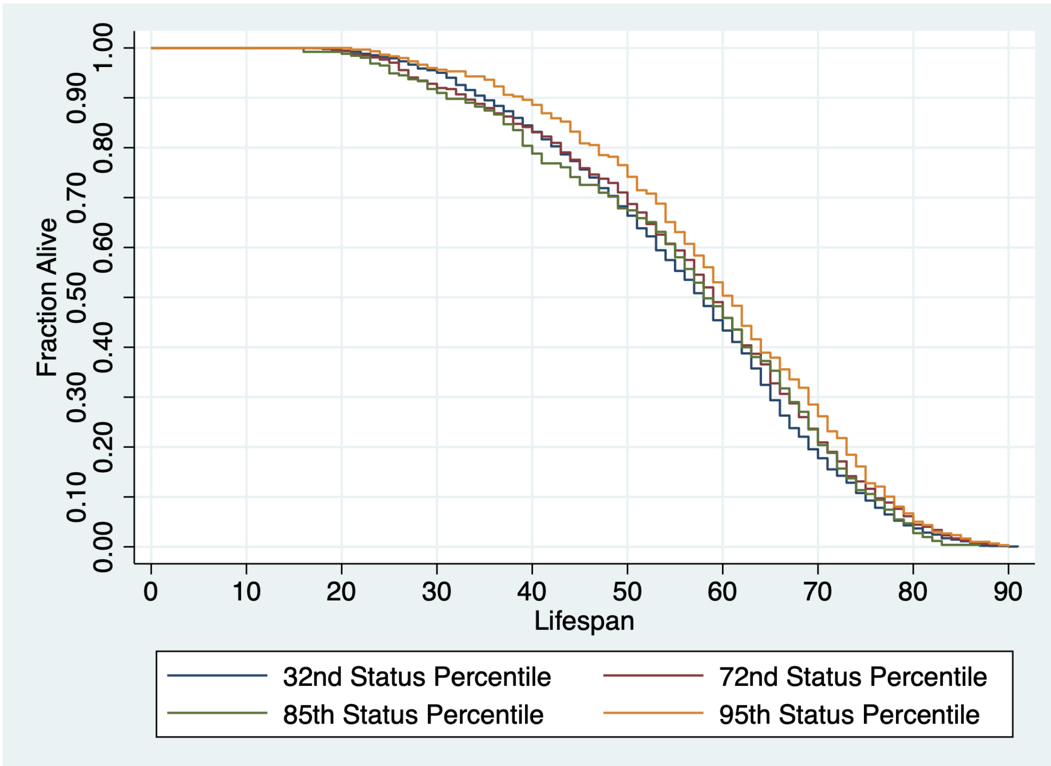


Notes: Scatter plot showing average percentile status by lifespan bin; limited sample n = 2,888.

Similar to the linear gradient, the health-status relationship exhibits more noise in the limited compared to the baseline sample. This pattern is due in part to the greater representation of high-status men within the group with the lowest lifespan, presumably because early deaths were more likely to be recorded for high-status compared to low-status men, even if the latter was more common. The unreconstituted sample reflects this selective omission of low-status men who died relatively early.

Figure A.7 shows Kaplan-Meier survival estimates for the limited sample.

Figure A.7: Survival by Status: Limited Sample

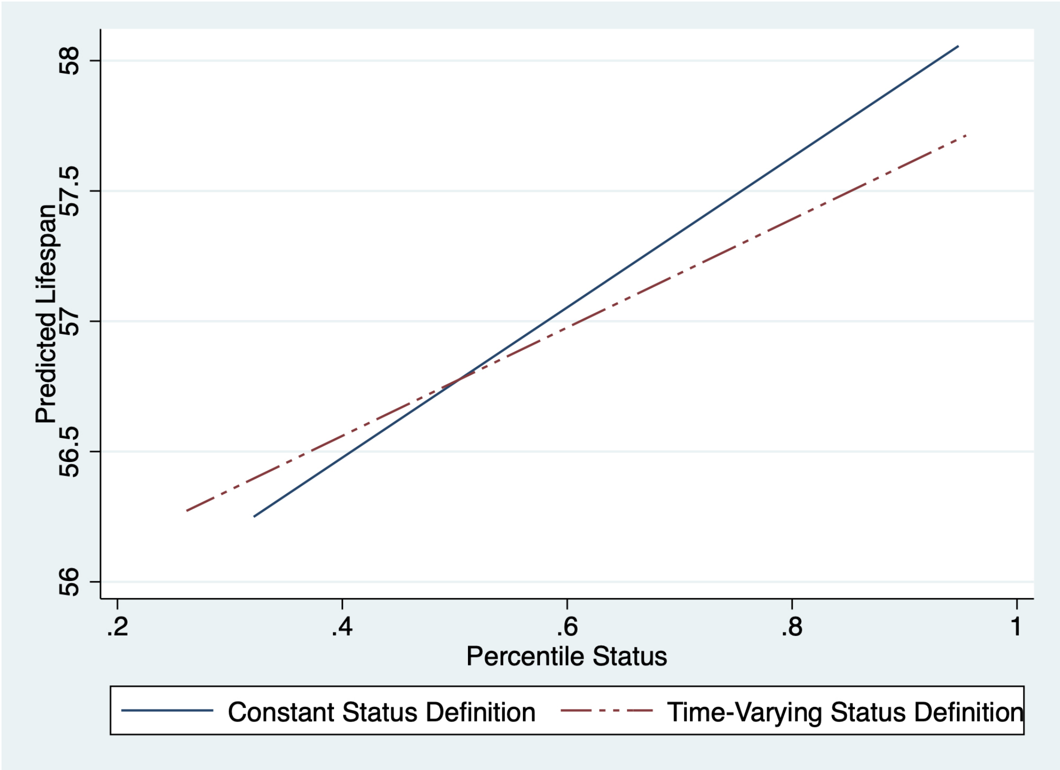


Notes: Kaplan-Meier survival estimates by status; n = 2,888.

Survival analysis for the smaller sample shown in Figure A.7 confirms the finding in the text that the chance of survival is higher for men of higher status. At the same time, the magnitude of survival differences for men of different status levels is smaller than in the full sample. For example, the difference in the probability of survival to 50 between highest and lowest status men is now 8 percentage points (75% versus 67%), in contrast to 14% for the baseline sample. This parallels our findings regarding the steepness of the Preston curve above.

Next, we examine the role of a time-varying status definition for the Preston curve in the limited sample.

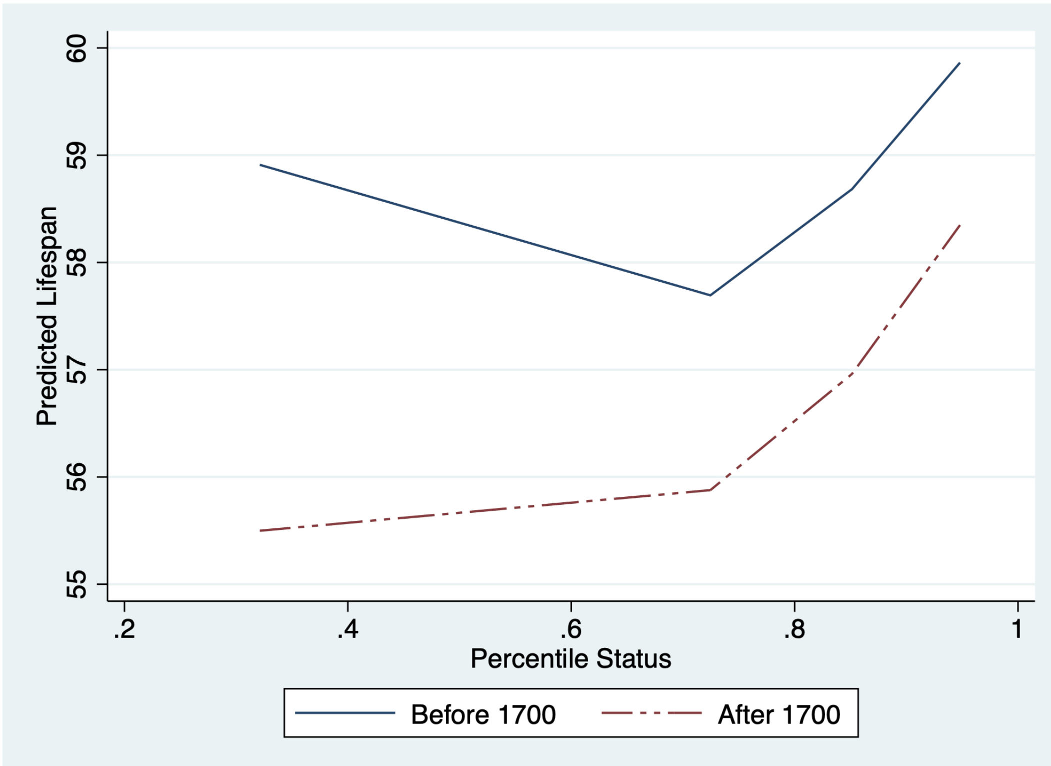
Figure A.8: Time-Varying Status Distribution: Limited Sample



Notes: Linear fit for constant and time-varying status definition (birth cohorts before and after 1700); baseline sample.

Figure A.8 shows that in the limited sample, the steepness of the linear status-health gradient is somewhat lower once we adopt a time-varying status distribution. In contrast, for the larger baseline sample our findings were similar for constant versus time-varying status definition. That the rescaling of status by subperiod matters somewhat more now is due in part to the lower number of observations in the limited sample. Figure A.9 shows results for the status-health relationship over time for the limited sample.

Figure A.9: Status-Health Relationship over Time: Limited Sample

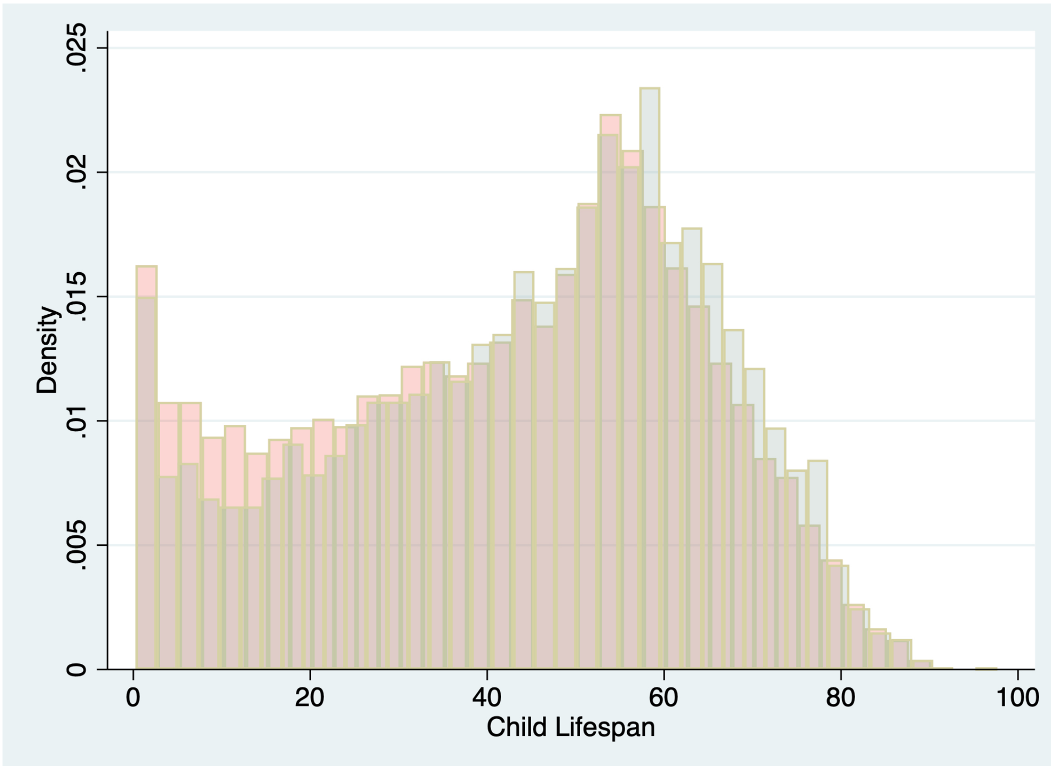


Notes: Polynomial fit for birth cohorts before 1700 (n = 845) and birth cohorts 1700 and later (n = 1,971).

Figure A.9 shows evidence for a systematic positive health-status gradient for the later subperiod. In the earlier subperiod, the sample men with the lowest status have on average a lifespan of close to 59 years, higher in fact than men around the 80th percentile. The reason for this is likely unrecorded vital statistics of many poor men with relatively low lifespan. Also notice that the health-status relationship for the earlier subperiod is located fully above the relationship for the later subperiod. Because the earlier subperiod includes the Ming-Qing transition with relatively high levels of mortality, this also points to sample selection. Men in the limited sample during the earlier subperiod are positively selected in terms of status and lifespan, with vitals statistics included for low status men who disproportionately lived a longer life.

Figure A.10 shows the distribution of male lifespans, including vitals statistics for sons before marriage.

Figure A.10: Lifespan of Sons: Baseline versus Limited Sample



Notes: Red series is baseline, including data reconstitution; green series is based on unreconstituted data.

We see that using the unreconstituted data shifts the lifespan distribution of sons to the right. This pattern is in line with earlier findings that data is disproportionately incomplete for shorter lifespans and lower status individuals (Telford 1990). Also note that the strikingly low infant mortality (compared to estimates of the time period) even in the reconstituted data is the primary reason that we focus on adult lifespan for assessing the health-status gradient.

In the following we examine the extent to which son mortality varies with the social status of their fathers in the limited sample. Table A.3 presents results for mortality rates at different ages. The very low infant mortality rates (death within first 12 months of birth) are expected, given changing norms over the study period about when a son and his premature death would be recorded in the vital statistics. Accordingly, account for omitted infant mortality in interpreting survival to age 5, 10, or 15, and in comparing these Chinese patterns of child survival to those available for historical European populations.

We see that also in the limited sample, mortality rates tend to be lower the higher is social status of the father. At the same time, we see that the difference in mortality rates between the baseline and limited samples varies with status. For example, the under-ten mortality

Table A.3: Mortality Rates of Sons: Limited Sample

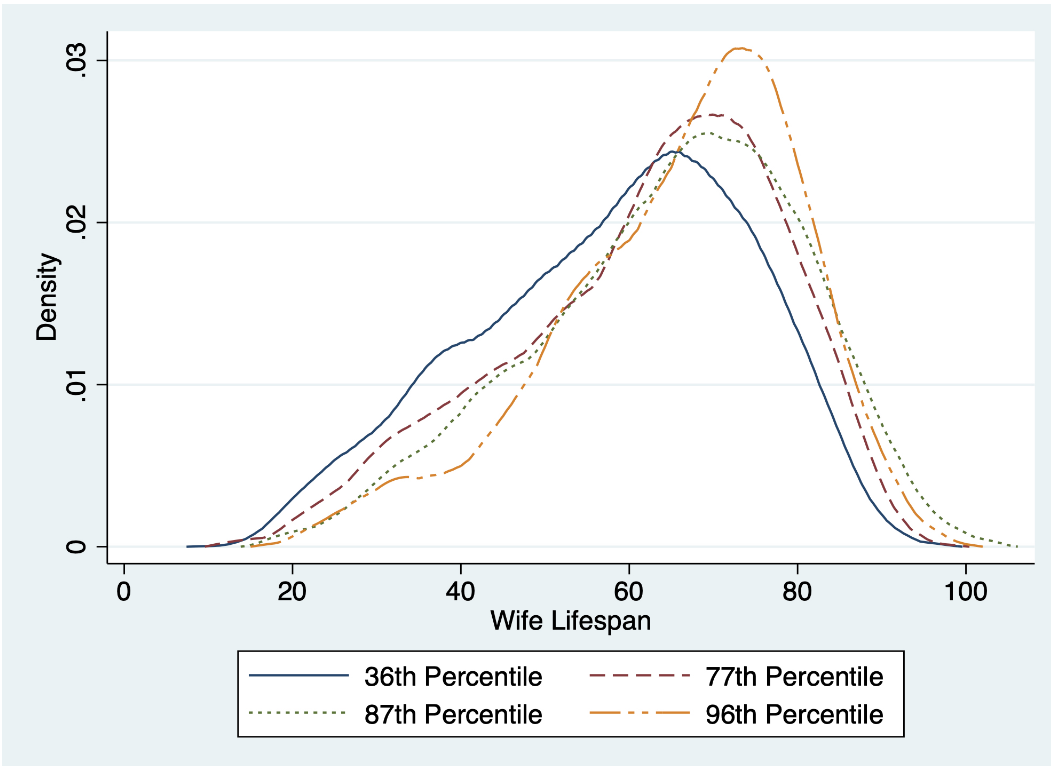
	N	One Year or less (12 months)	Five Years or less	Ten Years or less	Fifteen Years or less
32th Percentile Status	4,116	0.68	5.56	9.57	13.27
72th Percentile Status	1,088	1.01	6.71	11.21	14.51
85th Percentile Status	576	0.69	6.08	8.85	11.11
95th Percentile Status	710	0.14	2.82	4.51	6.06
All	6,490	0.68	5.50	9.23	12.43

Notes: Mortality rates given in percent. The number of sons for which father status information in the limited sample is available is n = 6,490.

rate for the highest status sons is 4.51% in the limited and 4.56% in the baseline sample, a one percent higher rate for the baseline sample. In contrast, under-ten mortality rates for the lowest status sons are 12.4% in the baseline and 9.57% in the limited sample, which is a thirty percent difference. This pattern is what one would expect if the unreconstituted data undersamples lower-status observations with relatively short lifespans.

Figure A.11 shows the distribution of wife lifespan by the status of the husband for the limited sample (using only complete vital statistics for the wives).

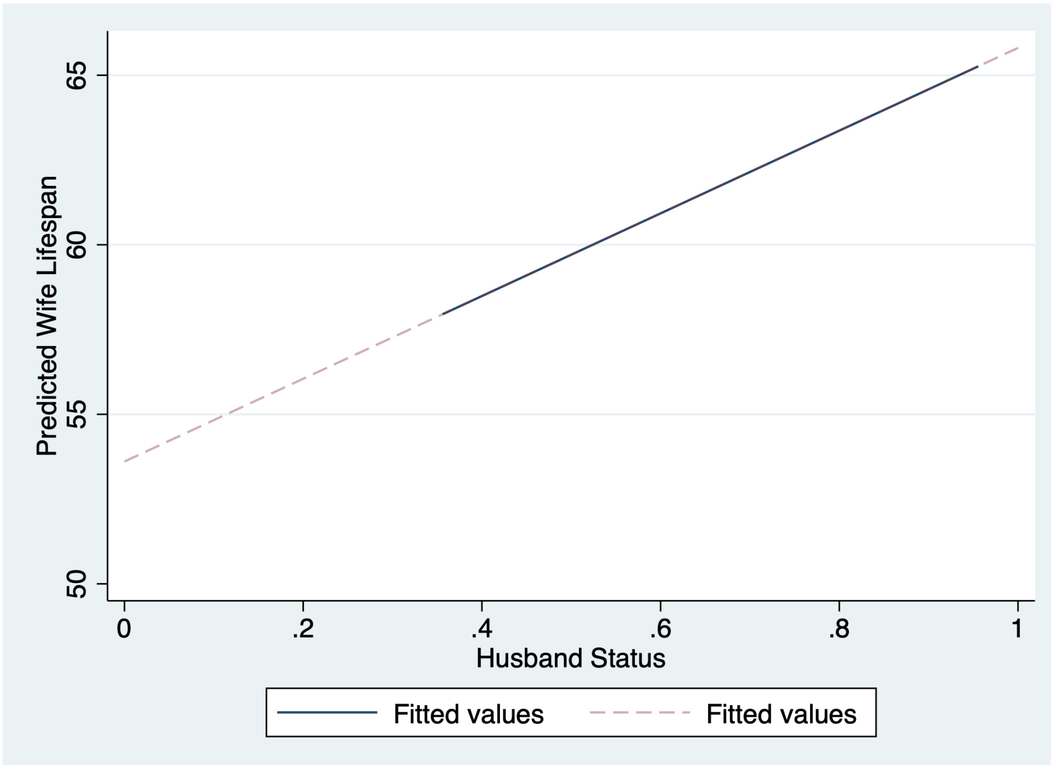
Figure A.11: Wife Lifespan by Husband Status: Limited Sample



Notes: Distribution of wives' lifespans for wives married to men belonging to seven Tongcheng clans born 1775 or before; n = 9,406.

We see that there is evidence for a strong female health-status gradient. In fact, while using the limited sample reduces the steepness of the male gradient, this is not the case for wife lifespan. This is confirmed by the linear relationship between husband status and wife lifespan shown in Figure A.12. Not only do wives generally live longer than husbands, but their lifespans are also comparably strongly related to husband status as in the baseline sample, implying more than a ten-year difference in lifespan between the lowest and highest status women.

Figure A.12: Linear Relationship between Husband Status and Wife Lifespan: Limited Sample



Notes: Distribution of male childrens' lifespan of married men belonging to seven Tongcheng clans born 1775 or before; n = 9,406.

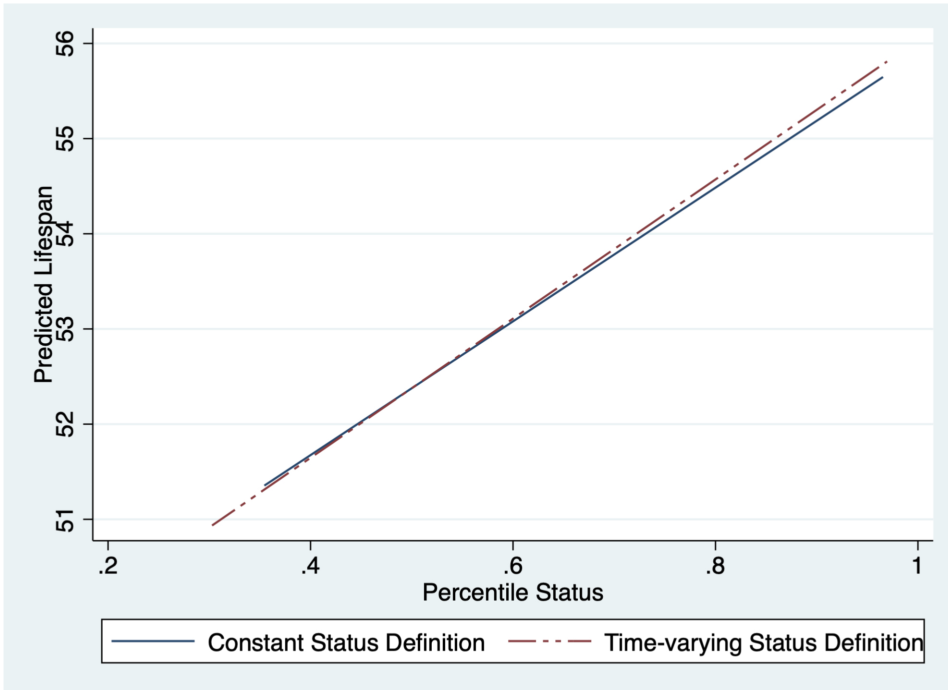
C Robustness

C.1 Time-Varying Status Distribution

This section explores the role of changes in the ability of the men to achieve high status over the sample period. Existing research notes that at least in the later part of the Qing dynasty it became harder to achieve high status levels. One reason was that despite high rates of population growth, the number of official positions remained about the same. Thus, while the average percentile status for the entire sample period is equal to 0.5 by construction (see Table 1), the average status of cohorts born before 1700 is 0.55, while cohorts born after 1700 have only an average percentile status of 0.47.

To see what role this changing status distribution plays in our Preston curve findings, we calculate the mens' percentile ranks separately for birth cohorts before and after 1700. This

Figure A.13: Health-Status Gradient and Changes in Mean Status



Notes: Linear fit for constant and time-varying status definition (birth cohorts before and after 1700); baseline sample.

controls for the fact that status became harder to obtain after 1700 because by construction, the average percentile status rank is 0.5 for each subperiod (born before and after 1700). Figure A.13 compares the Preston curve for a constant and time-varying definition of status. Figure A.13 indicates that results that account for status becoming harder to obtain are similar to findings with a constant status distribution (as in Figure 3 above). We conclude that the role of changes in the status distribution over the relatively long sample period do not play a major role for our findings.