Supercentenarian and remarkable age records exhibit patterns indicative of clerical errors and pension fraud

Saul Justin Newman*[1]

[1]Biological Data Science Institute, Australian National University

*Correspondence to: saul.newman@anu.edu.au

Abstract

The observation of individuals attaining remarkable ages, and their concentration into geographic sub-regions or 'blue zones', has generated considerable scientific interest. Proposed drivers of remarkable longevity include high vegetable intake, strong social connections, and genetic markers. Here, we reveal new predictors of remarkable longevity and 'supercentenarian' status. In the United States supercentenarian status is predicted by the absence of vital registration. In the UK, Italy, Japan, and France remarkable longevity is instead predicted by regional poverty, old-age poverty, material deprivation, low incomes, high crime rates, a remote region of birth, worse health, and fewer 90+ year old people. In addition, supercentenarian birthdates are concentrated on the first of the month and days divisible by five: patterns indicative of widespread fraud and error. As such, relative poverty and missing vital documents constitute unexpected predictors of centenarian and supercentenarian status, and support a primary role of fraud and error in generating remarkable human age records.

Introduction

The concentration of remarkable-aged individuals, within geographic regions or 'blue zones' [1] or within databases of people exceeding extreme age thresholds [2,3], has stimulated diverse efforts to understand factors driving survival patterns in these populations [4,5]. Populations within remarkable-age databases and 'blue zone' regions have been subject to extensive analysis of lifestyle patterns [5–8], social connections [4,9], biomarkers [10,11] and genomic variants [12], under the assumption that these represent potential drivers behind the attainment of remarkable age.

However, alternative explanations for the distribution of remarkable age records appear to have been overlooked or downplayed. Previous work has noted the potential of bad data [13], population illiteracy [14] or population heterogeneity [15] to explain remarkable age patterns. More recent investigations revealed a potential role of errors [16–19], and potential operator biases [20] in generating old-age survival patterns and data. In turn, these findings prompted a response with potentially disruptive implications: that, under such models, the majority if not all remarkable age records may be errors [21].

Here, we explore this possibility by linking civil registration rates and socioeconomic data to per-capita rates of remarkable age attainment, using data from every known centenarian (individuals aged 100 or over), semisupercentenarian (SSCs; aged 105 or over), and supercentenarian (aged 110 or over) from the USA, France, Japan, the United Kingdom, and Italy (Fig 1).

These data reveal that remarkable age attainment is predicted by regional indicators of error and fraud including greater poverty, higher illiteracy, higher crime rates, worse population health, greater levels of material deprivation, shorter average lifespans, fewer old people, and the absence of birth certificates. In addition, French and Italian historical data indicate that supercentenarians are not likely to be born into longer-lived cohorts, but are born into undifferentiated or shorter-lived populations relative to their contemporary national averages. Supercentenarian birthdates also exhibit 'age heaping' distributional patterns that are strongly indicative of manufactured birth data. Finally, fewer than 15% of exhaustively validated supercentenarians are associated with either a birth certificate or a death certificate, even in populations with over 95% death certificate coverage.

As such, these findings suggest that extreme age data are largely a result of vital statistics errors and patterns of fraud, raising serious questions about the validity of an extensive body of research based on the remarkable reported ages of populations and individuals.

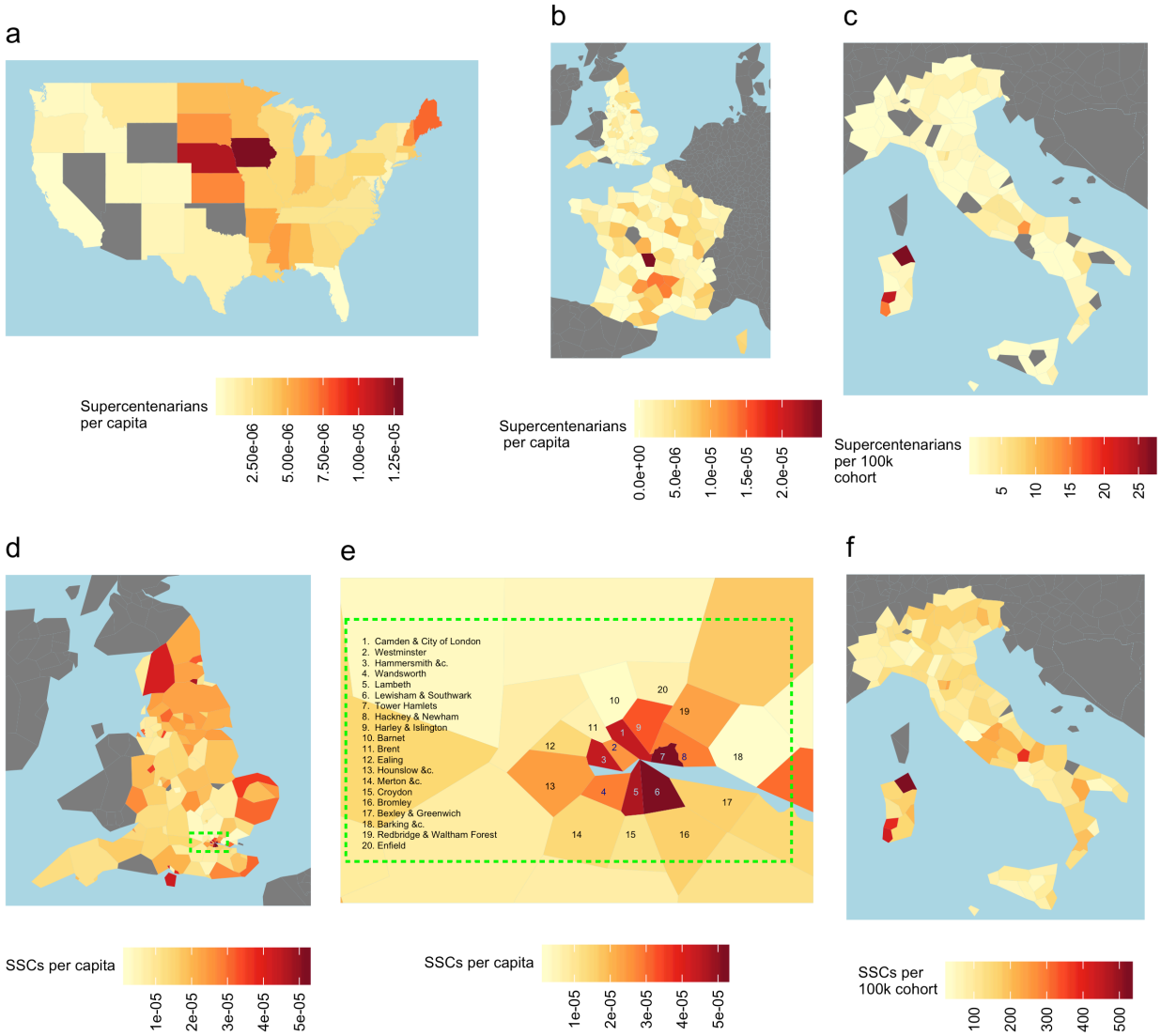


Figure 1. Regional distributions for the density of remarkable longevity. The large majority of SSCs are concentrated in a few countries, each exhibiting large regional variation in density of remarkable longevity records. Most supercentenarians are concentrated in the USA (a), with large numbers in France and the UK (b), and Italy (c; life table estimated rates). Within these countries, there exists marked variation in the density of remarkable age records: for example, variation in SSC density per capita across the UK (d) includes a 19-fold difference in SSC abundance within London (e) between Tower Hamlets (region 7; most) and Barnet (region 10; least), with similar variation in SSC density across the Italian provinces (f).

70

Results

After removing countries with incomplete data or inadequate spatial resolution, and omitting individuals with unknown birth locations, this analysis included all known individuals over age 100 in Japan and Italy, all known SSCs and supercentenarians in the UK, the 99.4% of known GRG supercentenarians in the USA with documented birth locations (IDL data have had such data removed or omitted), and all 175 supercentenarians in France with documented birth locations. In total, over 81% of the total global supercentenarian population were located to their region of birth.

Between the 1880 and 1900 census, a period covering 79% of US supercentenarian births, the US population increased by 150% and average life expectancy by twenty per cent [22,23]. The introduction of complete-coverage vital registration in the USA coincided with this rapid increase in lifespan and population size, and was expected to result in a large increase in the number of supercentenarian records per capita.

Instead, the introduction of state-wide birth certification coincides with a sharp reduction in the number of supercentenarians born in each state. In total, 82% of the GRG supercentenarian records from the USA predate state-wide birth certification. Forty-two states achieved complete birth certificate coverage during the survey period. When these states transition to state-wide birth registration, the number of supercentenarians falls by 80% per year overall (Fig 2a) and 69% per capita (Fig 2b) when adjusted relative to c.1900 state population sizes. The observed drop in supercentenarian number after birth registration remained after right-censoring the GRG

data by as much as 10 years to allow for the delayed or incomplete reporting of recent deaths (S1 Code).

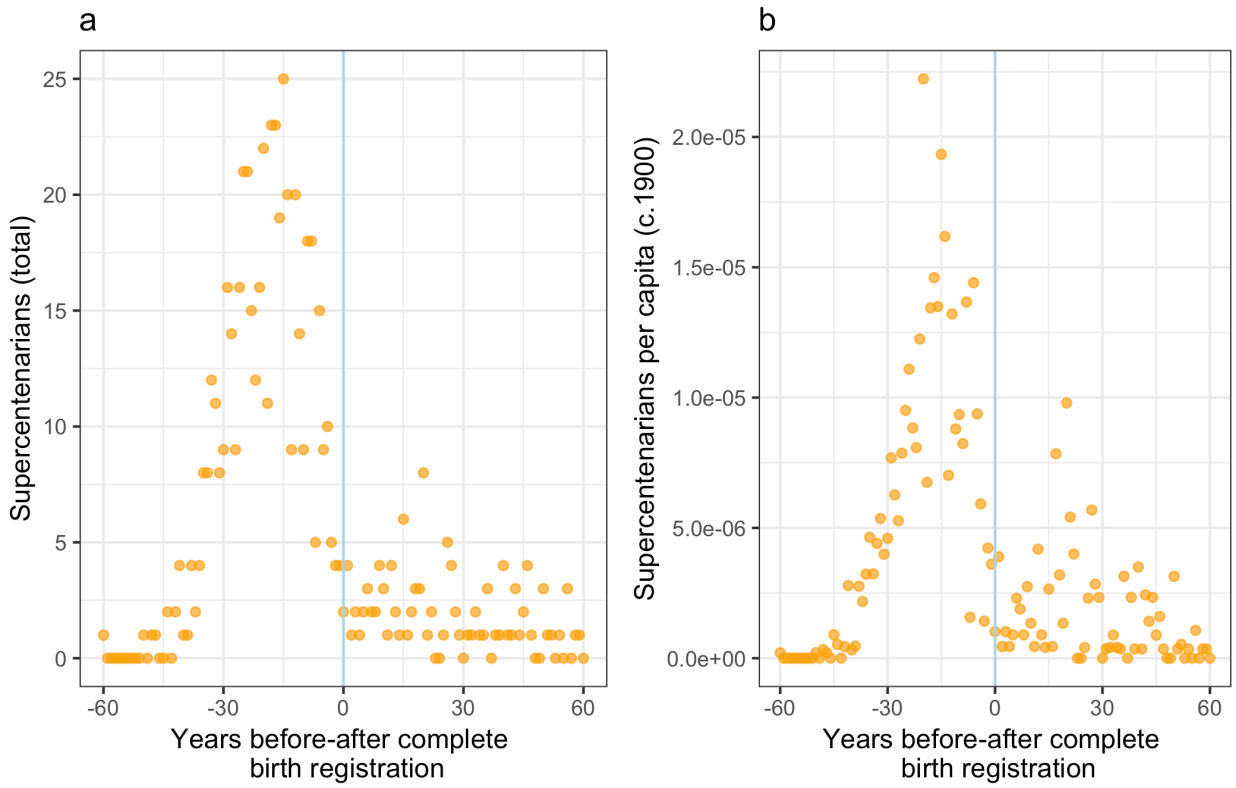


Figure 2. Number and per capita rate of attaining supercentenarian status across US states, relative to the introduction of complete-area birth registration. Despite the combined effects of rapid population growth and increasing life expectancy during this period the total number of US supercentenarians in the GRG database (a) falls dramatically after birth certificates achieve state-wide coverage (vertical blue line). This trend remains after adjusting for total population size c.1900 (b) within each state.

In countries with more complete birth documentation, high poverty rates were the best predictor remarkable age records. This interaction was unexpectedly positive, with increased poverty

predicting a higher density of centenarians per capita in Japan (Fig S1c), SSCs per capita in the UK, and supercentenarians per capita in both France and the UK (Fig 3; S2 Code). Old-age specific measures of poverty, available in France and the UK, increased the strength of this relationship. Higher rates of old-age poverty are linked to higher densities of remarkably old people: the amount of old-age poverty alone predicts up to seventy per cent of the variation in extreme longevity (Fig 3; S2 Code).

When aggregated into the larger NUTS-3 regions containing at least one SSC, the level of income deprivation in older people or IDOP, an indicator of the fraction of people aged 60+ suffering poverty and income stress, predicts around 40% of the variation in SSC abundance per capita across regions in the UK ($r = 0.42$; $p = 0.0000009$; 127 regions; Fig 3a). That is, throughout the UK higher income deprivation rates in older people predict higher numbers of people surviving past age 105. The accuracy of this basic poverty model improves markedly when predicting the number of SSCs as a percentage of all 90+ year old residents, rather than the number of SSCs per capita ($r = 0.70$; $p < 1x10^{-15}$; Fig 3d). As such, the highest concentration of remarkable lifespans in the UK arise in London's east end boroughs near Stepney and the Isle of Dogs, followed by urban Manchester, Tyneside, and Liverpool: the UK 'blue zones' (Fig 1; Table S1).

As in the UK data, the density of French supercentenarians is also predicted by higher poverty rates in the population age 75 and over ($r = 0.42$; $p = 0.0003$). However, this number was markedly reduced by the three overseas departments (Guadeloupe, Martinique and La Réunion) included in the regression which were marked outliers for both poverty and supercentenarian

rates (S2 Code). If these three outliers were trimmed, correlations between over-75 poverty rates and supercentenarian abundance increased (r = 0.58; p = 3e-07; Fig 3e; S2 Code). Again, like the UK data, these estimates also improved for predictions of the percentage of 90+ year-olds who also exceeded age 110. Across the 66 regions of France with sufficient data, the poverty rate over age 75 predicted half of the variation in supercentenarian abundance per 90+ year old (r=0.51, p= 0.00001; Fig. 3e).

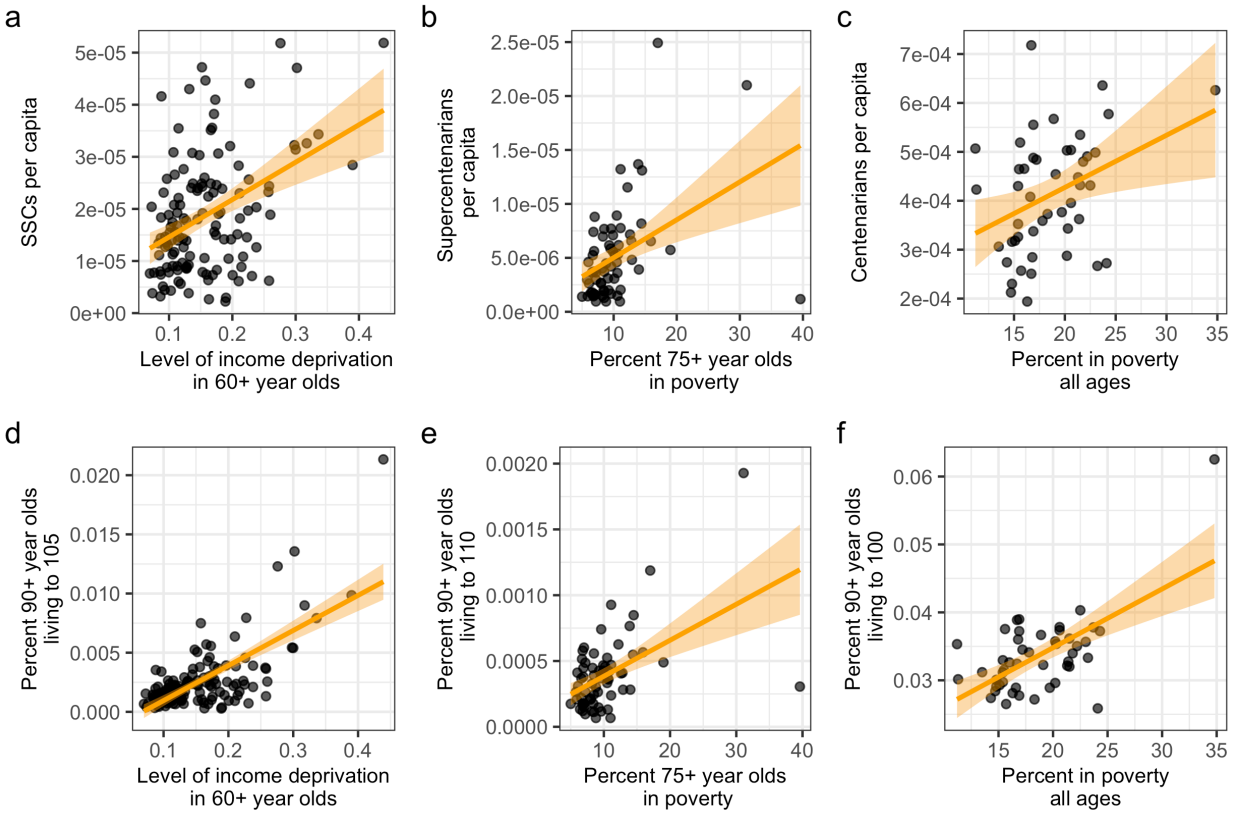


Figure 3. Old-age poverty and the density of remarkable lifespans. A higher percentage of people are centenarians, SSCs and supercentenarians in high-poverty and income-deprived regions of rich countries. Metrics of poverty and old-age poverty are positively correlated with the density of SSCs per capita in the United Kingdom (a; r = 0.42; p < 0.000001; 127 regions), supercentenarians in France (b; r = 0.42; p = 0.0004; 66 departments; r = 0.58 in the 63 mainland departments), and centenarians in Japan (c; r = 0.36;

p = 0.01; 47 prefectures). These relationships strengthen if density is instead measured by the fraction of 90+ year old people who exceed age 105 (d, UK; r = 0.71, p < 2e-16), age 110 (e, France; r = 0.51; p = 0.00001), or age 100 (f, Japan; r = 0.62, p = 0.000004).

In Japan, poverty rates in the general population were also positively correlated with the density of remarkable lifespans (r = 0.36; p = 0.01; Fig 3c). Again, this interaction strengthens when poverty rates are used to predict the fraction of 90+ year old people living past age 100 (r = 0.62; p = 4e-06; Fig 3f). While old-age specific estimates of poverty in Japan appear to be unavailable, although language barriers made this uncertain, such estimates are forthcoming. It can be predicted that these estimates of old-age poverty may provide an even better predictor of centenarian abundance than overall poverty rates.

In addition to the unexpected relationship with poverty rates, the number of centenarians in Japan is also negatively correlated with income per capita (r = -0.44, p=0.001), the minimum wage (r = -0.64; p = 1e-09), and the Japanese financial strength index (r = -0.70; p = 3e-08) across all 47 Japanese prefectures. Prefectures that spend more money on old-age welfare per capita, a disincentive for welfare fraud, also produce fewer centenarians per capita (r = -0.49; p= 0.0004). These factors share latent drivers and are highly autocorrelated: prediction models for centenarians per capita based solely on poverty (R^2 = 0.37; p= 0.0001; S2 Code) approach the accuracy of linear mixed models containing all available socioeconomic variables (R^2 = 0.43; p = 3e-06) yet have a lower Akaike's information Criteria (S2 Code).

In the UK, which had the most abundant granular data available at the NUTS-3 regional level, socioeconomic indicators collectively predicted half of the regional variation in SSC density

when fitted as interactive effects in a linear mixed model (Fig 4; adjusted R^2 = 0.0.39; p = 6.09e-05; S2 Code). Under a one-way analysis of variance, the best predictors of SSC density per capita were the Indices of Multiple Deprivation or IMD, a multidimensional indicator of area-level hardship and poverty (Fig 4a; F value = 20.3; p = 2e-05), the IDOP indicator of old-age poverty (Fig 5b; F value = 16.5; p = 9e-05), and purchase-power standard adjusted gross domestic product or PPS-adjusted GDP (F value = 10.3; p = 0.002; S2 Code). When predicting the number of SSCs per 90+ year old across the UK, this socioeconomic model captures 80% of the variance in SSC abundance (adjusted R^2 = 0.80; p < 2.2e-16; S2 Code), again largely through the effect of IMD (F value = 61.8; p = 6e-12), IDOP (F value = 326.4; p < 2.2e-16), and the interaction between IDOP and the UK crime index (F value = 29.7; p = 4e-07), all of which predict higher SSC density under worse conditions.

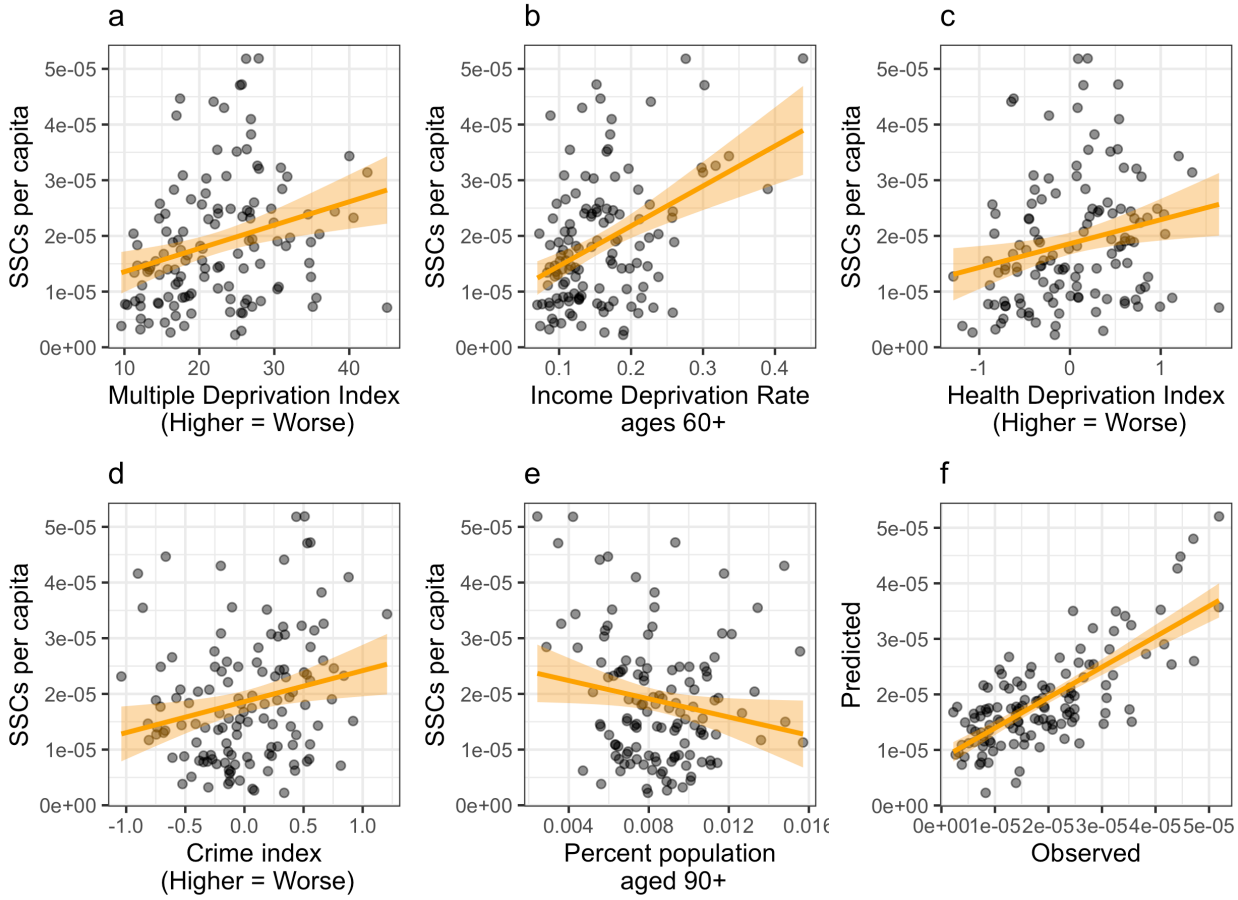


Figure 4. Social hardship and the distribution of 105+ year old people in the United Kingdom. Across 128 regions containing at least one semisupercentenarian, the number of SSCs per capita is highest in regions where: (a) people are more deprived overall ($r = 0.28$; $p = 0.001$), (b) people over age 60 are more income-deprived ($r = 0.42$; $p < 0.000001$), (c) multidimensional health indices are worse ($r = 0.23$; $p = 0.01$), and (d) crime rates are higher ($r = 0.22$; $p = 0.01$). In contrast, regions with a higher fraction of people surviving past 90 years of age (e), have significantly fewer SSCs per capita ($r = -0.18$, $p = 0.04$). A simple linear mixed model (f) fit to these five variables, PPS-adjusted GDP, and employment rates (S2 Code), accurately predicts the regional density of SSCs across the UK ($r = 0.74$; adjusted $R^2 = 0.39$; $p = 1e\text{-}06$).

Fitted as interactive effects in a linear mixed model, general and old-age poverty rates predicted 46% of the variance in supercentenarian density across France (overseas territories included, adjusted $R^2 = 0.44$; p = 1e-08; S2 Code). When PPS-adjusted GDP per capita, unemployment and murder rates per capita were also included as fixed effects, these inputs collectively captured 46% of the variance in supercentenarian density across France, an increase caused by interactions of PPS-adjusted GDP with both overall and old-age poverty rates (adjusted $R^2 = 0.56$; p = 1e-08; S2 Code).

Direct measurements of provincial poverty rate were not available for Italy at the NUTS-3 regional level. Instead, the attainment of remarkable age in Italy is predicted by worse early- and mid-life health. While survival to age 55 is positively correlated with life expectancy at all ages at all ages from 60 to 95, higher early- and mid-life survival are inversely correlated with mortality rates after age 95 (Fig 5a). Cohort survival to age 55 is increasingly negatively correlated with survival to ages 100 (Fig 5b), 105 (Fig 5c) and 110 years (Fig 5d), and with life expectancy at age 100 (r = -0.4; p=0.00001; S2 Code). That is, autocorrelation between age-specific survival breaks down and inverts at advanced ages, such that better survival to mid-life is linked to worse survival in advanced age. Contrary to expectations yet again, both a lower probability of survival to age 55, and higher probability of death at age 55, are linked to a higher density of remarkable age records (S2 Code).

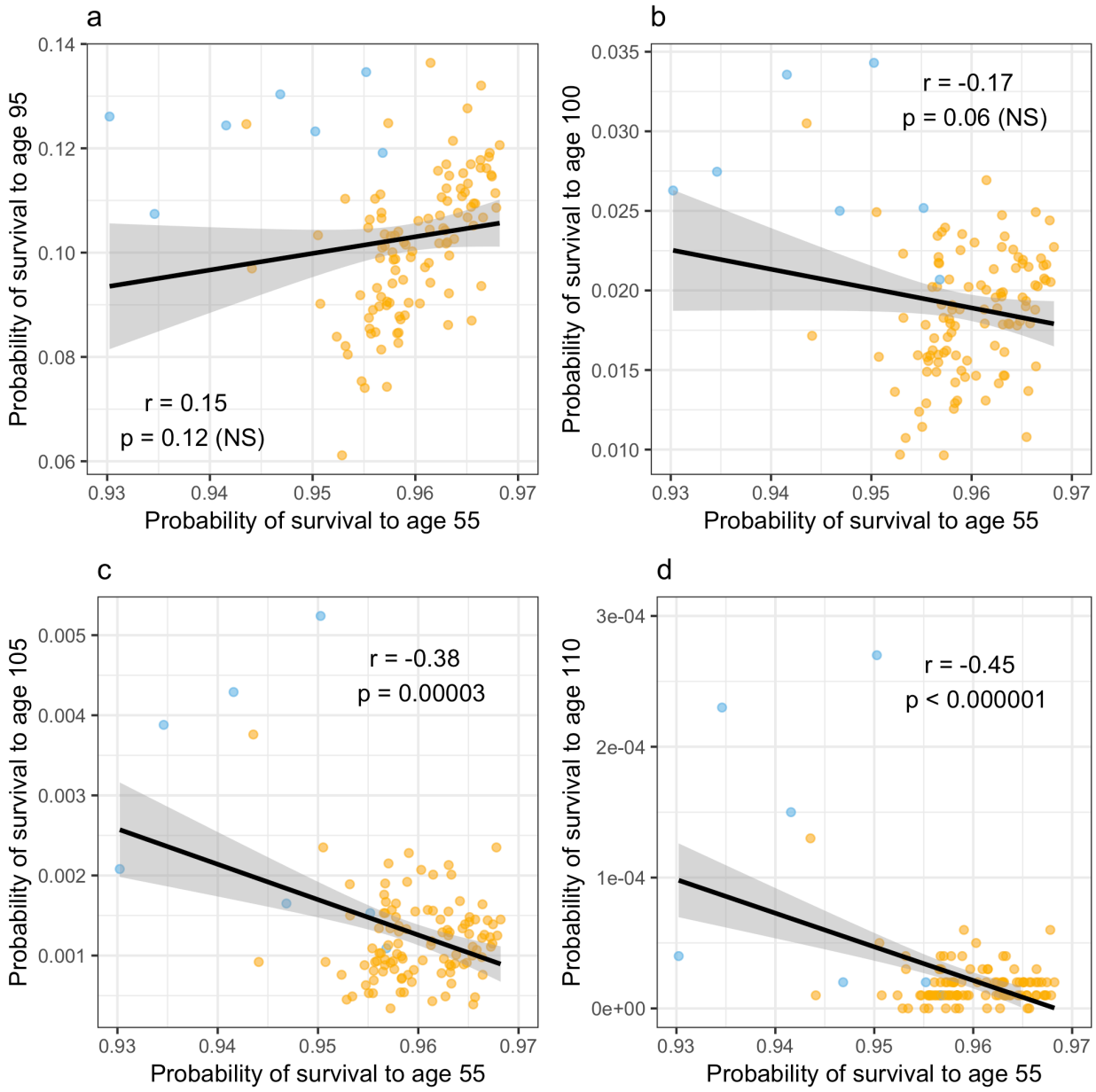


Figure 5. Relationship between mid-life and late-life survival across Italian provinces. Across Italian provinces (points), probabilities of survival in mid-life are positively correlated with the probability of survival at older ages until around age 95 (a; r = 0.15; p=0.1; N=116). However, this relationship inverts at advances ages: better mid-life and early-life probabilities of survival, and higher average longevity, are linked to significantly lower probabilities of survival at 100 years (b), 105 years (c), or 110 years (d) of age. Sardinian provinces shown in blue.

Population sizes, GDP per capita, PPS-adjusted GDP per capita and employment rates were available at a sufficiently granular level across Italy for basic analysis. Individuals across Italy were more likely to attain supercentenarian ages if their province has a worse economy (Fig S2a-c), higher unemployment rates (Fig S2d-f), and fewer people over the age of 90 (Fig S2g-I; S2 Code). However, a linear mixed model with cohort life expectancy to age 55 had a significantly lower Akaike's information criteria, compared to a linear mixed model containing these basic economic indicators as interactive effects (S2 Code).

As in Italy, French historical data did not reveal the expected positive relationship between life expectancy and supercentenarian status. There are 143 French supercentenarians whose region of birth had a corresponding local and national estimate of life expectancy at birth. For these individuals, cohort life expectancy for the region and year of birth was lower than, but not significantly different from, the contemporary national average ($p = 0.52$; $N = 143$; Fig S3). That is, supercentenarians were not, on average, born into regions with either significantly longer or shorter than expected life expectancy across metropolitan France. It seems unusual that modern economic conditions and poverty rates are predictive of reaching age 110, yet life expectancy at birth is not (Fig S3).

While informative of general trends, these models mask large regional anomalies in the pattern of remarkable age records. Several of these anomalies, particularly those regions with the very highest number or density of extreme age records, require some comment.

For example, Scotland and Northern Ireland have a combined modern population of seven million people, yet produced only three known SSCs. In contrast, the 24-fold smaller population of Tower Hamlets has produced fifteen SSCs: the most SSCs per capita in the UK (Table S1). However, Tower Hamlets also has the highest poverty rate, highest child poverty rate, the shortest disability-free life expectancy, the highest income inequality, [24], and the worst index of multiple deprivation [25] of all 32 London boroughs. Of all 317 local authority districts in the UK, Tower Hamlets has the single most income-deprived population of older people [25]. Of these 317 local area districts, Tower Hamlets also has the smallest percentage of people aged 90 and over [25]. This is a notable discrepancy.

Adjacent to Tower Hamlets, the borough of Southwark & Lewisham ranks second for SSCs per capita, and first for SSCs overall (N= 31) across the 175 NUTS-3 regions of the UK (Table S1). Southwark & Lewisham is also the eighth most income-deprived district for older people in the UK, and has the fifth-fewest 90+ year old people per capita [25]. Outside of London, Tyneside and Greater Manchester are strongly represented. For example, Manchester produced 18 SSCs overall (equal sixth) and ranks 14th for SSCs per capita (Table S1), yet is the third most income-deprived district for older people in the UK, and has the highest crime index, third-worst population health index, fourth-worst index of multiple deprivation, and sixth smallest percentage of 90+ year old people of any region (Table S1). These rankings are not a recent shift: Manchester is the second most-persistently deprived of the 317 local-authority districts in the UK [25].

As in the UK, French regions illustrate a concordance between regions with the highest poverty and the regions with the most remarkable age records. The mainland region of Creuse has the highest per capita rank of supercentenarians under the IDL rankings (Table S2), potentially due to the combined effects of having a 60% reduction in population size since 1901, the fourth-highest old-age poverty rate, the 16th worst poverty rate overall, and the fourth-lowest GDP per capita (Table S2).

Guadeloupe and Martinique rank equal second for total supercentenarians after Paris, and second and third for supercentenarians per capita, with at least eight supercentenarians each (Table S2). Martinique has the second-highest poverty rate, both overall (29%) and for people aged 75 and over (31%), in all of the 101 NUTS-3 coded provinces. While not monitored by Eurostat, third-ranked Guadeloupe has a 24% unemployment rate [26]. Of the Eurostat regions only Réunion has higher poverty rates, with 39% of citizens falling below the poverty line. Again, these rankings seem inconsistent with the general drivers of population health.

French supercentenarians are over-represented in the overseas departments, former colonial holdings, and Corsica (which is included in metropolitan France; Table S2): regions that historically constitute some of the most neglected, least well-documented, and shortest-lived administrative regions of France. As a result, many of these regions are absent from the above models due to absent or insufficient population data and reporting.

At the first reliable estimate of population size in 1950, overseas departments and colonial holdings contained around 1.7% of French citizens. However, at least 11% (N=16) of the French

supercentenarians in the GRG database originate there: a 6.5-fold over-representation. This number increases when integrating deidentified IDL data, which only includes regions monitored by Eurostat (Réunion, Guadeloupe, and Martinique), to establish a minimum numbers of supercentenarians born in each region. Under these estimates, the overseas and former colonial regions of France contain a minimum of twenty-four (15.5%) of the total 155 supercentenarians with known birth locations. Guadeloupe and Martinique each contain eight supercentenarians, and French Algeria contains four, Saint Barthélemy, Réunion, French Guiana, and New Caledonia at least one each.

If this minimum count across the IDL and GRG databases is used, the overseas and colonial holdings of France contain eight times as many supercentenarians per capita (1.5 per 100,000), and more supercentenarians overall (N=24), than the region of Île-de-France (0.16 per 100,000; N=19). This is despite Île-de-France earning more than double the income per capita, being the longest-lived region in mainland France, and containing seven times as many citizens c.1900 when these supercentenarians were born.

Similar anomalies occurred in Japan and Italy. Of 114 total regions, the Italian province of Olbia-Tempio ranked as the best province for survival to ages 100,105 and 110, yet somehow was also the seventh-worst ranked province for survival to age 55, and according to Eurostat had the eighth-fewest residents surviving over the age of 90 (Table S3). The first and second ranked Japanese prefectures for centenarians per capita, Shimane and Kochi, had the worst and second-worst regional economic rankings, while extensive anomalies in third-ranked ‘blue zone’ Okinawa are detailed below (Table S4).

Overall patterns observed in the IDL and GRG database also provided indications of widespread error. Contemporary US births have a near-uniform distribution of births with minimal deviation from random sampling (Fig S4a), driven by the aseasonal and approximately equal distribution of days of the month (*e.g.* the 1st or 2nd of each month) throughout the year. Even after the widespread uptake of induced births and surgical births, which avoid weekends and public holidays, birthdays generally varying by less than two per cent across different days of the month. In contrast with this uniform distribution, supercentenarians in the GRG database are 142% more likely to be born on the first day of the month and 1.2-fold as likely to be born on a day that is divisible by five (Fig S4b). The number of supercentenarians born on the first day of the month is 150% higher than on the preceding calendar day (S1 Code). Given the near-complete absence of caesarean sections in these population, this age-heaping pattern can be explained if a large percentage of people are non-randomly choosing their birthday.

The first day of the month was not over-represented amongst the IDL data, and the 5-day enrichment was less pronounced (Fig S4c). This was initially difficult to reconcile, given these databases overlap considerably and ideally contain the same set of individuals. However, this differential appears to be a result of dates of birth being removed for 48% (N=797) of the IDL supercentenarians.

These patterns can be explained because most of the signal for age heaping in the GRG data arises in Japanese and US birthdates (S2 Code): Japanese supercentenarians were 2.77 times more likely and US supercentenarians were 1.57 times more likely to be born on the first day of

335 the month. The IDL does not document Japan and unusually, despite comprehensive US birth dates being known and available, every US supercentenarian in the IDL database has had their day and month of birth removed. No other country received a similar treatment. With these dates excised from the database, evidence of heaping in the IDL data is reduced (Fig S4c; S1 Code).

340

Discussion

Basic economic and social indicators in the modern economy, such as GDP per capita and poverty rates, provide adequate predictors of the distribution of extreme age records. Despite constraints on model construction and accuracy, such as unavoidable differences in per capita adjustments, these basic models approached reasonable accuracy. However, the direction of these interactions is the opposite of rational expectations.

Diverse social and economic indicators normally linked to worse health outcomes, such as income deprivation, poverty, and high unemployment, are all positively associated with a higher probability of reaching an extreme age. These factors are linked to a lower probability of survival and worse health outcomes at every age below 90, for every population included in this study.

However, these factors exhibit a consistent positive association with extreme longevity. In the UK, which contains the only national data with sufficiently granular regional health measures, even poor health itself is positively associated with attaining a remarkable lifespan (Fig 4c).

Viewed in isolation such a question may, perhaps, be explained away by reference to unknown lifestyle factors. However, these findings should be considered in the context of other diverse and incongruous patterns observed in extreme old age studies.

Indicators of error and fraud in national data

Data used in this study raise simple questions as to why basic socioeconomic indicators of poor health, and positive correlates of crime and government neglect, are linked to higher per capita numbers of remarkable longevity. For example, the UK has produced 1075 SSCs overall, and Italy 3,638 SSCs overall, across an approximately equivalent timeframe [27]. However, Italy is a historically smaller, poorer, less well-educated, and shorter-lived country. In 1900 the UK had eight million more inhabitants than Italy, a 1.22-fold larger population [28]. Citizens of the UK also enjoyed 2.5 times the GDP per capita, earned 3.5 times higher wages in real terms, had 1.25 times lower income inequality, received 2.2 times the average education (with just 5.3 years of schooling), were four times less likely to be murdered, were 3.8cm taller, and lived 5.3 years longer on average than people in Italy [28]. Given these indicators and the long history of birth records in both countries, it is difficult to reconcile why the healthier, wealthier, better-educated, taller, and longer-lived population of the UK produced roughly a quarter as many SSCs per capita. One explanation is that remarkable age records result, not from better health or greater longevity, but from the historical accumulation of illiteracy-driven errors and the modern dynamics of poverty-driven fraud.

Relative to the global average, states containing remarkable age records generally constitute rich, literate, long-lived and well-documented populations, usually with an extensive history of vital statistics documentation. As a result, the existence of widespread errors and pension fraud is often assumed to be unlikely or impossible in these countries. However, such national advantages are no guarantee of data quality.

High-quality universal registration systems often contain undetected high-frequency errors. Contrary to previous assertions that "*Japan has…among the highest quality data for the oldest-old*" [29], a 2010 investigation of Japanese records revealed that 238,000 centenarians were actually missing or dead[30]. The Japanese Ministry of Health and Welfare [31,32] now estimate there were 43,882 Japanese centenarians alive in 2010: an 82% reduction, and a notable contrast to the idea that "*Japanese demographic data have always been considered extremely reliable*" [33].

Similar instances have occurred elsewhere. In 1997 Italy discovered it was paying 30,000 pensions to dead people [30]. In the USA, a recent analysis cross-checking census and death records found at least 17% of centenarians and 35% of 109+ year-olds were actually errors [16,34]. In 2011, just one of several Greek social insurance institutes was caught paying the pensions of 1,473 dead people who had 'survived' past the age of 90. A subsequent 2012 investigation by the Greek labor ministry was triggered by the "*unusually high number of 9,000 Greek centenarians drawing old-age benefits*" [35], a notable figure given the 2011 Greek census found only 2,488 living centenarians [36]. Assuming the census contains no type I errors, which is unlikely given the high rate of active pension fraud, at least 72% of Greek centenarians had been collecting their pensions whilst dead.

Despite the Greek labor ministry paying a fraction of all Greek pensions, its investigation revealed over 200,000 pensions were being paid to fraudulent claimants including 'blind' taxi drivers and dead people [35]. An estimated two per cent of Greeks were engaged in benefits

fraud at the time of the 'blue zone' surveys, and thousands of these dead pensions were still being paid in 2013.

These examples refute claims that the suggested existence of widespread pension fraud or errors in databases of SSCs and supercentenarians is "*not credible*" [21]. Unlike academics seeking to generate old-age databases, governments have both a direct financial incentive to detect pension fraud, and the considerable resources required to do so: and yet, developed-world governments routinely fail to detect thousands of cases of document-based pension fraud.

Indicators of error and fraud in blue zones

Results presented in this study may reflect a general neglect of error processes as a potential generative factor in remarkable age records, and the omission of evidence from national statistical bodies. This potential for disregarding important context and national statistical data may be most evident when considering the case of 'blue zones': proposed regions of remarkable longevity [37].

The 'blue zone' of Okinawa has the highest number of centenarians per 90-99-year-old of any Japanese prefecture and remains world-famous for remarkable longevity[37]. However, according to the statistics bureau of Japan, Okinawa also has the fewest senior citizens per capita, the highest murder rate per capita, the worst over-65 dependency ratio, the second-lowest median income, and the highest unemployment rate of some 47 Japanese prefectures [32]. Despite prior claims of dietary benefits based on vegetable and sweet potato consumption [37],

Okinawa has the single lowest per capita intake of sweet potato, at 64% of the Japanese average intake. Okinawa also has the single lowest consumption per capita of fruit, vegetables, seafood, taro, shellfish, root vegetables, pickled vegetables, and oily fish such as sardines and yellowtail [32]. Okinawa has the second-highest per capita intake of beer, the fourth-highest alcohol consumption, the most 'flophouses', the most 'shotgun' weddings, the highest per-capita intake of Kentucky Fried Chicken [32], and according to USDA estimates Okinawans consume an average 14 cans of spam per year each [38]. Okinawa has a 36% child poverty rate, 15% higher any other prefecture [39]. Mortality rates in Okinawa 'cross over' after age 50, such that older individuals and cohorts have age-specific mortality rates far below the national average [33]: a pattern indicative of unreliable data and misreported ages [13]. Okinawa also has the second-lowest minimum wage (by one yen), the lowest household savings, the highest percentage of over-65s on income assistance, the highest poverty rate [39], and the worst average body mass index of all 47 prefectures [32]. These rankings have not changed substantially since the blue zone surveys and do not represent a recent sudden shift away from traditional lifestyles. Again, it seems unusual that so few of these issues have been raised by an extensive body of demographers, epidemiologists, and public health scientists familiar with the 'blue zones' concept.

There are other well-known drivers of error in the Japanese vital registration system. Birth and marriage records in Japan are not generated by a central bureaucracy, but instead are hand-recorded by family members as 'Koseki' documents, which are then filed in local town halls and government offices. This combination of citizen self-reporting and government filing allows the

propagation of errors without requiring fraud. In Okinawa, this broad potential for error generation has also been compounded by a different class of error processes.

The large-scale US bombing and invasion of Okinawa involved the destruction of entire cities and towns, obliterating around 90% of the Koseki birth and death records [33] with almost universal losses outside of Miyako and the Yaeyama archipelago [40]. Post-war Okinawans subsequently requested replacement documents, using dates recalled from memory in different calendars [40], from a US-led military government that largely spoke no Japanese. The number of these replacement Koseki documents issued, a proxy measure of American bombing and shelling intensity, predicts 79% of the variation in centenarian status in Okinawa [33].

Like the 'blue zone' islands of Sardinia and Ikaria, Okinawa represents deprived regions of rich, high-welfare states. These regions may have higher social connections, and arguably may have had higher vegetable intakes in the past, but they also rank amongst the least educated, poorest, highest-crime and least healthy regions of their respective countries.

These patterns were reflected in this study's findings on the 'blue zone' provinces of Sardinia, which during the 'blue zone' surveys had the highest murder rate in Italy [41]. The primary 'blue zone' province of Ogliastra has the single lowest survival rate to age 55 and the highest unemployment rate of 116 regions in Italy (Table S3), while potential issues in the Greek data may be clear from the previous discussion.

The two remaining blue zones, Loma Linda and the Nicoya Peninsula, are considered exceptional due to their high average longevity rather than the presence of extreme outliers for longevity. As such, these claims are not relevant to assessments of supercentenarian status, yet their analysis also raises serious questions and broader issues in extreme longevity research. For example, Loma Linda is a Californian suburb containing just 23,000 people, designated as a 'blue zone' because of an estimated average lifespan of 86 years for females and 83 years for males. This average lifespan is matched or exceeded by the 125 million citizens of Japan, the seven million citizens of Hong Kong, and the seven and a half million citizens of Singapore [42]. When assessed independently by the Centers for Disease Control (CDC) the five small-area survey tracts covering Loma Linda instead have an average life expectancy of 76 to 81 years [43]: the 27th to 75th percentiles of US life expectancy (Fig S5). At best, the independent CDC estimates rank Loma Linda as the 16,102nd most long-lived neighborhood in the USA (Fig S5; S1 code).

Loma Linda is not a standard census tract but a custom-selected region within a larger geographic area, the largest US county of San Bernadino, with an average lifespan of 78 years [43,44]. The Nicoya peninsula in Costa Rica, where independent estimates are currently not available, is also a non-standard region cut from several independent census units of moderate life expectancy by proponents of the 'blue zone' concept. The first 'blue zone' was described by drawing circles on a map with a blue pen [45] across two standard, independently surveyed regions of Italy with the lowest and sixth-lowest probability of survival to age 55 [46]. As such, it seems somewhat debatable that these regions should be regarded valid outliers for human longevity.

Indicators of error and fraud in health studies

Like anomalous population-scale patterns, indicators of poverty and fraud and contra-indications of health are regularly ignored or downplayed in studies of extreme age. For example, smoking rates of *e.g.* 17-50% and illiteracy rates of 50-80% are often observed in samples of the oldest-old [7,8]. Surveying the 'blue zone' of Ikaria, Chrysohoou *et al.* observed that the oldest-old have: a below-median wage in over 95-98% of cases, moderate to high alcohol consumption (5.1-8.0 L/ year), a 10% illiteracy rate, an average 7.4 years of education, and a 99% rate of smoking in men [4].

In the Tokyo study of exceptional longevity 15.4% of centenarians were current smokers [47]. However, according to the national statistics bureau of Japan, above the age of 80 only 3.9% of Japanese women (78% of the Tokyo sample were women) and 19.3% of men are smokers [48]. Tokyo centenarians smoke at around twice the rate that could be expected in a younger 80+ year old cohort with an identical sex ratio. Likewise, 80% of the 'exceptional' health-status centenarians in Tokyo drank alcohol every day, followed by 49% of the 'normal' and less than 40% of the 'frail or fragile' centenarians, which resulted in "*a* [significant] *positive relationship between drinking habits and functional status*" [47]. In contrast, in the general population only 2.8% of women and 23% of men aged 80+ drink every day [48]. In men aged 60-69, the heaviest-drinking cohort in Japan, daily drinking rates peak at 36.7%: less than half the rate of the exceptionally healthy centenarians [48]. Tokyo centenarians drink at higher rates than any other age group, and smoke at rates equal to a population 45 years younger [48].

In the USA only 8.4% of general population over the age of 65 smoke, and in Europe 4.1% population over the age of 75 smoke: figures that continue to fall with age due to two-fold higher mortality rates in smokers [49,50]. However, in the US and Europe individuals over the age of 100 smoke and drink at much higher rates [4]: in one US centenarian study [6] 60% of people over age of 95 were former smokers, compared to just 25% of individuals over the age of 65 in the broader USA.

Anomalies where harmful health behaviors become more prevalent with age is common in centenarian studies. For example, comparisons of lifestyle factors in centenarians to earlier surveys of the same cohort [6], revealed that centenarians have a similar or worse body mass index, and worse rates of physical activity, smoking, and alcohol consumption, than younger representatives of their birth cohort. Rajpathak *et al.* concluded that these similar or worse lifestyle indicators meant centenarians were representative of the baseline population, documented by the NHANES I survey, from which they were drawn [6]. However, this comparison was made between a population who were on average 97.5 years old [6], and individuals who were aged 35 years younger.

Longitudinal follow-up surveys of the NHANES I comparison cohort reveal a linear decrease in the number of smokers and drinkers with age until age 85 that is typical of almost all populations. This reduction is caused by increased all-cause mortality, at all ages below 95, in individuals who smoke and drink [51,52]. After just 9.2 years, over 33% of people who were current smokers in the NHANES I cohort had quit smoking, 8.9% remained current smokers, and

13% were either current or former smokers: the rest were deceased. Therefore, the observation of 60% smoking rates in centenarians [6], who are older representatives of the same cohort, presents some difficulty.

The increasing frequency of past harmful behavior with age also occurs in longitudinal studies of the oldest-old, suggesting these patterns do not result from ascertainment biases or population differences. For example, in the Leisure World Cohort Study of over 11,000 retirees 78% of men and 72% of women in the initial 71-year-old population were moderate to heavy drinkers [53]. Individuals who drank daily throughout the 23-year follow-up period of this study had a significantly lower mortality risk, even those drinking 2-4 times over the CDC clinical threshold for 'excessive' drinking [54]. In contrast with normal clinical patterns, abstaining from alcohol significantly increased mortality risk while drinking at the most dangerous levels, 28 or more alcoholic beverages a week, was associated with an unexplained 9-16% reduction in the risk of death [54].

It is unclear why clinically excessive drinkers or daily smokers should survive at equal or higher rates, and increase in population frequency at extreme ages, unless these lifestyle factors are positively correlated with committing fraud or having an incorrect age.

The frequency of birth and death certification, indicators that are linked to data quality but not health outcomes, are suggestive of the latter. In the benchmark New England Centenarian Study, only 30% of enrolled centenarians have an official birth document of any description [55]. This rate of documentation is lower than the background population, with every state in New England

achieving state-wide birth certificate coverage by 1897 [22]. For the remaining 70% of centenarians without birth documents, age validation was carried out using US census data, as a best-case scenario, or documents including "*military certificates, an old passport, school report card, family bible, and baptismal or other church certificate*" [56]. Twenty-two years after the New England study used US census data for validation and study enrolment, and after the publication of extensive research findings, at least 17% of centenarians in the US census data were discovered to be errors [16,34].

Birth certificates are also rare in databases of remarkable age records. Just 19% of all supercentenarians and 20% of the 'exhaustively' validated supercentenarians are listed as having either an original or copied birth certificate by the IDL [57]. Overall only 6.6% of supercentenarians have an original birth certificate, and 74% of cases have no reported birth documents of any kind: not even parish register data is recorded. None of the 797 SSCs from the USA is listed as having a birth certificate, and 41% of these cases have the possibility of a birth certificate explicitly ruled out [57].

These high rates of missing data may be due in part to low reporting or incomplete monitoring in some countries: death and birth certificates may exist, but have not been entered into the database. However, a low rate of reporting does not explain the absence of birth certificates in countries with near-complete reporting of evidence. The 241 French supercentenarians in the IDL all have the field code for birth documents completed, but not one supercentenarian has a birth certificate (S1 Code). Likewise, the IDL explicitly lists the evidence available in support 98% of remarkable-age claimants in the UK (S1 Code). Of the 1116 oldest people in the UK, all

590 but three were born after the introduction of compulsory, nation-wide birth certificates in 1864. However, only eighteen (1.6%) have a birth certificate. Furthermore, these rates do not increase markedly in the 1,587 'exhaustively' validated UK, French and US-born SSCs: in total only 24 have a birth certificate listed in support of their case, and all of these documents are re-issued copies.

595

While birth certificates were produced more than a century in the past, death certificates are issued by modern bureaucracies: over 96% of listed SSCs died after the year 1990, during a period of unprecedented death certificate coverage (S1 Code). However, like birth certificates, death certificates are completely absent or severely under-represented relative to their respective

600 birth cohort amongst individuals in remarkable age databases. This inconsistency is difficult to explain.

The first follow-up survey of the population-representative NHANES I cohort, representative of the US population, found 3.8% of decedent men and 5.7% of decedent women did not have

605 death certificates and remained alive on paper while actually dead [51]. While around 95% of the general US population are issued a death certificate, only seven of the 504 supercentenarians dying in the USA is listed as having a death certificate by the IDL: an over 70-fold lower rate of death certification. Despite a self-described exhaustive validation effort, just 1.4% of SSCs and supercentenarians have a death certificate in the USA.

610

The USA is not an isolated case. Across the IDL database only 15% of supercentenarians and 8% of SSCs are listed as having death certificates (S1 Code). These rates are often lower in

countries with more-complete and longer-term death certification histories [58]. For example, of the 9386 SSCs dying in France, just one has a death certificate [57]. Only 89 of the 1184 SSCs dying in the UK have an original or copied death certificate listed [57]. Both of these countries and have maintained over 90% death registration rates for several decades, and are now approaching 100% death certification rates [58]. It is therefore striking that, despite death certificates being issued to well over 90% of individuals, a substantial majority of validated remarkable-age individuals in these countries do not seem to have a death certificate.

Indicators of error and fraud in individuals

The absence of basic birth and death certification and the high prevalence of counter-indicators of health and longevity stand in contrast to the large number of listed cases of extreme longevity. In theory each such case is individually assessed and validated, based on the compilation of documents and the judgement of demographers. Assessing the role of opinion during case validation, and corresponding potential for bias, is therefore of marked importance.

Individual case studies often highlight the role of personal judgement, and the potential for both conscious and unconscious bias, during age validation. For example, Jiroemon Kimura, the world's oldest man, is widely considered to be a valid supercentenarian case. However, Kimura has at least three wedding dates to the same wife [59], has three dates of graduation from the same school [59], was conscripted to the same military three times in four years [59] despite the mandatory conscription period being three years long [60], and has at least two birthdays [59].

For the first 20 years of his life all of Kimura's birthdates and school records are actually recorded for a different name, Kinjiro Miyake, whose connection to Kimura is attested to by a hand-written note from a Korean mail and telephone company [59] rather than any official document. The evidence for Kimura's case validation was initially compiled and vetted by the relatives who sought to promote his case [59]. Under interview, Kimura then explained one of his extra birthdays in a way that was "*not feasible*" [59], and Gondo *et al.* concluded the birth date had been deliberately forged [59]. However, Gondo *et al.* resolved the case validation by assuming any conflicting official records were mistakes and, from the diverse birth, wedding, conscription, and graduation dates, selecting the dates they felt were accurate. Multiple names, multiple weddings, and forged birthdates notwithstanding, the study concluded that "*no critical discordances were discovered*" [59] and the case is considered valid [57,61].

The validity of the Kimura case has been accepted under the assumption that age discrepancies can be discarded through the qualitative judgement of demographers. Reliance on such qualitative judgements during case validation is considered acceptable conduct. In addition, concerns surrounding the validity of ages are often met with the response that biographical inconsistencies, detected during interview by a demographer, will result in cases being removed from the record.

However, this sentiment can be difficult to reconcile with observed practice. Former smoker and occasional drinker Adele Dunlap, who "*ate anything she wanted*" and "*never went out jogging or anything*" [62] was validated by the GRG and IDL as the oldest woman in the USA, despite Dunlap consistently maintaining under interview that she was a decade younger: if "*asked how it*

felt to be 113, Dunlap… looked her questioner in the eye and answered: 'I'm 104'" [62]. Despite consistently maintaining until her death that her age was incorrect, Dunlap remains validated as a supercentenarian on the basis of documentary evidence. This documentary evidence has since lowered her age by two years in just one [61] of the two [57] major supercentenarian databases.

A reliance on this type of opinion, where qualitative judgements are employed to shape public perceptions of authenticity, seems to be widely considered satisfactory. This seems particularly the case when explaining the otherwise anomalous health habits of supercentenarians. For example, Maier *et al.* issued a contradictory statement that Jeanne Calment smoked both one and two cigarettes a day for an entire century, followed by the justification that this counter-indication of health could be explained because she "*possibly did not inhale at all*" [63]. It was likewise observed that, from age 20 to age 117, the then-oldest man in the world Christian Mortensen smoked "*mainly a pipe and later on cigars, but almost never cigarettes… he had also chewed tobacco…but never inhaled*" [63]. Why two people would voluntarily choose to smoke for an accumulated 190 years, yet never inhale, was never explained.

Such behavior is not atypical. At least three of the ten oldest women drank every day, two smoked every day, and four are of unknown smoking status, while Jeanne Calment smoked daily, drank daily, and ate around a kilogram of chocolate a week. Of the five oldest men ever recorded, Kimura and Mortensen are detailed above, Emiliano Del Toro (3rd) smoked for 76 years, Mathew Beard (4th) was busted for drink-driving at age 90, and Walter Breuning (5th) smoked cigars until he was 108. The oldest man in the UK stated his secret to health as "*cigarettes, whiskey and wild, wild women*" [64], while the former oldest man in the USA started

every day with coffee and whiskey, drank during the day, ate ice-cream every night, and smoked from age 18 until his death. Like Calment and Mortensen, he also didn't inhale some 12 to 18 cigars a day [65].

These instances of poor lifestyle choices constitute a substantial fraction of all supercentenarian cases. As summarized by Coles, the typical supercentenarian lifestyle is characterized by "*heavy smoking, heavy drinking, or both, failure to exercise on a regular basis, and no conscious effort to eat nutritiously*" [66]. Instead of prompting skepticism, under the relatively safe assumption that smoking, drinking, poverty, lack of exercise, poor nutrition, and illiteracy should not enrich for remarkable longevity records, these anomalous contra-indications of survival are routinely ignored or downplayed. For example, the study by Chrysohoou *et al.* concluded that "*physical activity, dietary habits, smoking cessation, and midday naps*" predict extreme longevity in the Ikaria 'blue zone' [4]: a conclusion that questionably re-shapes past smoking status as a positive indicator of survival. Genetic factors that convey a collective immunity to cancer and the diverse sequelae of smoking, drinking, and not exercising are also frequently raised as an explanation for the lifestyles of the extremely old [6,12,67].

In contrast, it could be suggested that the abundance of poor lifestyle choices in the extreme old reflect high rates of undetected error. If this were the case, a large body of previous research linking higher old-age survival to, for example, higher drinking [47,54] and obesity rates [68,69] could be re-interpreted as the result of a positive correlation between poor lifestyle factors and 'junk' vital statistics data.

705

Type I error detection in extreme age databases

It seems incongruous that the discovery of thousands or hundreds of thousands of fake centenarians by the respective Japanese, Greek and Italian governments and US researchers, has
710 not resulted in any corresponding reduction in the size of supercentenarian databases. Instead, the number of validated supercentenarians increased smoothly across these fraud-discovery events. Perhaps the more limited resources of individuals compiling old-age records, using identical documents and similar techniques to government demographers, far exceeds the capacity of developed-world governments to detect identity fraud. Alternatively, perhaps,
715 supercentenarian databases remain riddled with error.

Data cleaning and error correction using documentary validation, as described in the Kimura case above, remains the main approach to combat age errors in remarkable longevity databases. However, data cleaning often produces the mistaken impression that the resulting 'validated'
720 data are largely free from error. Data in this study exhibit patterns consistent with a high frequency of type I errors: diverse positive correlates of crime, anomalous poor health indicators, age heaping, and over 20-fold higher rates of missing documentation than the general population. However, these populations were already subjected to extensive analysis and validation [27] and are widely considered high-quality 'clean' data [21].

725

The logic supporting these assumptions of data-cleanliness is informative. For example, post-validation errors in the Italian data were previously assumed to be minimal, on the basis of a

belief that the data were clean [27]. Subsequently, it was acknowledged that an unknown number of errors in these data could not be detected using documentary evidence, as "*Occasionally…a mistake will escape even a rigorous validation procedure*" [21]. Finally, it was proposed that the occurrence of such errors, which cannot be detected using documents, must be rare or "*essentially impossible*", because of the high quality of documents used to compile such data [21]. That is, type I errors are assumed to occur at low frequency on the basis of documentary evidence: documentary evidence that cannot detect the frequency of type I errors.

The opinion that such errors are rare might have been countered by another opinion: that a handwritten century-old database containing millions of entries, no independent biological validation, and an unknown type I error rate, might easily generate the few hundred annual errors required for a supercentenarian database [66]. Prior observations that the Italian state paid pensions to 30,000 deceased people [30], or that 82% of Japanese and 72% of Greek centenarians were illusory or dead [35,70], or suggest the viability of this explanation. However, such criticism would ignore a more fundamental problem.

Physical possession of valid documents is not an age guarantee. Consider a room containing 100 real Italian supercentenarians, each holding complete and validated documents of their age. One random supercentenarian is then exchanged for a younger sibling, who is handed their real and validated birth documents. How could an independent observer discriminate this type I substitution from the 99 other real cases, using only documents as evidence?

Such hypothetical errors cannot be excluded on the basis of document consistency: every document in the room is both real and validated. In addition, a real younger sibling is also likely to have sufficient biographic knowledge to pass an interview: this has occurred in several (subsequently discovered) cases including, for several years, the world's former oldest man. As such, any similar substitution error has the potential to indefinitely escape detection.

This 'Italian sibling' thought experiment illustrates why type I age-coding errors cannot be ruled out, or even necessarily measured, on the basis of documentary evidence. It also reveals how debates on the frequency of these errors are not driven by direct empirical measurements, but by inference and opinion.

This issue presents a substantial problem for remarkable-age databases, embodied in a deliberately provocative, if seemingly absurd, hypothesis:

Every 'supercentenarian' is an accidental or intentional identity thief, who owns real and validated 110+ year-old documents, and is passably good at deception.

This hypothesis cannot be invalidated by the further scrutiny of documents, or by models calibrated using document-informed ages [71,72]. Rather, invalidating this hypothesis requires a fundamental shift: it requires the measurement of biological ages from fundamental physical properties, such as amino acid chirality [73] or isotopic decay [74].

Until such document-independent validation of remarkable ages occurs, the type I error rate of remarkable human age samples will remain unknown, and the validity of ‘supercentenarian’ data in question.

775

Methods

The number and birthplace of all validated supercentenarians (individuals attaining 110 years of age) and semisupercentenarians (SSCs; individuals attaining 105 years of age) were downloaded from the Gerontology Research Group or GRG supercentenarian table [61], updated 2017, and the International Database on Longevity or IDL [57]. These data were aggregated by subnational units for birth locations, which were provided for the IDL data, and obtained through biographical research for the GRG data. Populations were excluded due to incomplete subnational birthplace records (<25% complete) or countries with an insufficient number of provinces to fit spatial regressions (<15 total provinces), leaving population data on SSCs and supercentenarians in the USA, France, and the United Kingdom (Fig 1).

To quantify the distribution of remarkable-aged individuals in Italy, province-specific quinquennial life tables were downloaded from the Italian Istituto Nazionale di Statistica Elders.Stat database [46] to obtain age-specific survivorship data (Fig 1c,f; S1 Code). Using cross-sectional data across Italian provinces, probabilities of survival (l_x) to ages 90-115, and life expectancy at age 100 were fit as dependent variables, and survival rates at age 55 and life expectancy at age 55 as independent variables, using simple linear regression (S1 Code).

While older ages were not available, extensive Japanese centenarian data were downloaded from the Japanese Ministry of Health, Labour, and Welfare [32] through the Statistics Japan portal [31] for all 47 prefectures (Fig S1), alongside data on prefectural income per capita (in 2011 yen), employment rates, age structure, survivorship, and a financial strength index, for 2010: the

most complete recent year available for these data (S1 Code; Fig S1). These data were also
800 linked to the most recent available prefecture-specific poverty rates [39].

Supercentenarians recorded in the GRG database and born in the USA were matched to the 1900 census counts for state and territory populations [22], and linked to the National Center for Health Statistics estimates for the timing of complete birth and death certificate coverage in each
805 US state and territory [75]. Both the number of supercentenarian births overall, and estimates of supercentenarians per capita, approximated by dividing supercentenarian number by state population size in the 1900 US census [22], were averaged across the USA and represented as discontinuity time series relative to the onset of complete-area birth registration (S1 Code).

810 To capture the geographic distribution of French supercentenarians, all 175 supercentenarians in the GRG database who were either born or deceased in France were linked to the smallest discoverable region of birth using biographical searches [61]. In addition, de-identified records in the IDL were already linked to birth locations encoded by the Nomenclature for Territorial Units level 3 codes (NUTS-3), which divide France into 101 regions [57]. These modern regions were
815 linked manually to their corresponding Savoyard-era department to obtain historic region-specific estimates of life expectancy at birth [76] for the birth year and location of all supercentenarians in metropolitan France. For each supercentenarian, life expectancy at birth was then measured relative to the contemporary average life expectancy of metropolitan France (S1 Code).

820

The number of total supercentenarians and SSCs born into Eurostat NUTS-3 coded regions, either documented for French and UK regions in the IDL or estimated for Italian regional cohorts by ISTAT, were linked to modern socioeconomic indicators available at this administrative level: total regional gross domestic product (GDP), GDP per capita, GDP per capita adjusted for purchase power scores (PPS), murder and employment rates per capita, and the number of 90+ year-olds, using the Eurostat regional database (S1 Code).

In the UK, additional data were obtained for the Index of Multiple Deprivation or IMD: a national metric used to indicate relative levels of deprivation, including income deprivation in people aged 60+, by the UK Office of National Statistics [25]. The IMD data are measured in 317 local authority districts, each of which is a subset of a single Eurostat NUTS-3 encoded region. To capture the relative degree of deprivation within the UK, the IMD and its component scores were averaged within each of the 175 NUTS-3 regions (S1 Code).

Similar estimates of deprivation were obtained for French NUTS-3 regions, by downloading the regional poverty rates [77] and poverty rates in the oldest available age group, ages 75 and over from the French National Institute of Statistics and Economic Studies INSEE [77].

To overcome the three orders of magnitude differences in population size across subnational geographic units, the number of centenarians, SSCs and supercentenarians were adjusted to per capita rates. However, the ‘correct’ adjustment for per capita rates of remarkable longevity is dependent on the *a priori* assumptions of their cause. For example, if the null hypothesis was that all supercentenarians are ‘real’, adjustment for birth cohort size 110+ years previously would be

a more correct method for best predicting the population density of supercentenarians. However, if the null hypothesis is that supercentenarians are more frequently modern-era pension frauds or clerical mistakes, per capita correction for a birth cohort 110 years in the past is of uncertain value for predicting modern events. In this latter case, the occurrence of supercentenarians would be better and more accurately predicted by correcting for modern population sizes.

The former 'historical per capita' adjustment was used whenever possible. Per capita rates of remarkable age attainment, calculated relative to the size of historical birth cohorts, were downloaded from the respective government statistical bureaus of Japan and Italy [32,46]. Due to the absence of birth certificates, USA supercentenarian data from the GRG [61] were corrected to per capita rates based on population data in the 1900 US census [22]. However, France and the UK were located into geographic units that have only existed since 2003. As a result, there were no data on historical population sizes available for these geographic units. It was therefore necessary to estimate per capita rates using modern population sizes surveyed at the NUTS-3 geographic level within France and the UK.

To address this unavoidable difference in per capita rate calculations the number of the centenarians, SSCs and supercentenarians were also corrected relative to the number of old-age residents in each modern geographic unit of Japan, the UK, and France (Fig 3d-f; S1 Code). This adjustment was less susceptible to large longitudinal shifts in population size, and better reflected the density of older people in modern geographic units after survival and migration processes. However, the insufficient granularity of birth cohorts within the UK, and the considerable

rearrangement of geographic units within France, remains an important constraint on the upper accuracy of these models.

Collective socioeconomic indicators obtained for each country were used to develop linear mixed models across all regions with a non-zero number cases, of centenarians in Japan, SSCs in Italy and the UK, and supercentenarians in the UK and France (S1 Code), to predict the regional per capita and per 90+ year old density of the oldest available populations in each country. Linear mixed models were fit using either the population poverty rate (UK, France, and Japan) or estimates of old-age poverty rates (percent in poverty over 75 in France, the IDOP index in the UK) as the single predictor variable, and the number of centenarians, SSCs and supercentenarians both per capita and per 90+ year old. These models were then extended by fitting, as interactive effects, basic socioeconomic indicators used as global indicators of health and deprivation available at a sufficient geographic level (S2 Code). Such models focused on capturing basic indicators, representing crime rates, health, and income, available at the NUTS-3 regional level in the EU and the prefectural level in Japan.

Where available, French supercentenarians were linked to regional estimates of life expectancy at birth, calculated quinquennially for each of the Savoyard-era departments of France into which they were born [76]. These local rates were then corrected relative to the contemporary French national average life expectancy at birth to yield the relative life expectancy at birth, in years [76]. For example, Jeanne Calment was born in the Alpes-Maritime department in 1875, when average life expectancy at birth was just 33.4 years and the contemporary national French average life expectancy was 37.8 years: a relative life expectancy of -4.4 years. These rates were

then used to estimate whether regional life expectancy at birth of French cohorts containing supercentenarians was significantly higher or lower than the French national average using a one sample t-test.

To explore the potential for age manufacture amongst remarkable age records, birthdate data were aggregated within the GRG and IDL databases. Enrichment for specific birth days is usually indicative of nonrandom age selection due to fraud, error, and clerical uncertainty. This check, however, is limited in that it cannot detect diverse sources of error, such as identity fraud or failed death registrations, which retain a representative distribution of birth days.

As population representative birthdates were unavailable within the target populations, the distribution of births was tabulated by days of the month to remove the often poorly-categorized or undocumented effects of birth seasonality. This distribution was compared to both modern birthdate distributions from seventy million births in the US, which suffer from increased distortion due to elective induced births and caesarean sections on certain dates, and to the distribution of birthdays under a uniform distribution of births.

To facilitate reproduction of these findings, all shareable data and code are available in a single structured file, with instructions and links for the non-shareable data, in S1 Data.

References

910 1. Buettner D, Skemp S. Blue Zones: Lessons From the World's Longest Lived. American Journal of Lifestyle Medicine. 2016. doi:10.1177/1559827616637066

2. Robine J-M, Gampe J, Vaupel J. IDL, the International Database on Longevity. Gerontology. 2005; 1–30.

3. Gerontology Research Group [Internet]. 2016 [cited 21 Oct 2016]. Available:
915 http://www.grg.org/SC/SCindex.html

4. Chrysohoou C, Panagiotakos DB, Siasos G, Zisimos K, Skoumas J, Pitsavos C, et al. Sociodemographic and lifestyle statistics of oldest old people (>80 years) living in ikaria island: The ikaria study. Cardiol Res Pract. 2011; doi:10.4061/2011/679187

5. Poulain M, Pes GM, Grasland C, Carru C, Ferrucci L, Baggio G, et al. Identification of a
920 geographic area characterized by extreme longevity in the Sardinia island: The AKEA study. Exp Gerontol. 2004; doi:10.1016/j.exger.2004.06.016

6. Rajpathak SN, Liu Y, Ben-David O, Reddy S, Atzmon G, Crandall J, et al. Lifestyle factors of people with exceptional longevity. Journal of the American Geriatrics Society. 2011. doi:10.1111/j.1532-5415.2011.03498.x

925 7. Afonso RM, Ribeiro O, Vaz Patto M, Loureiro M, Loureiro MJ, Castelo-Branco M, et al. Reaching 100 in the Countryside: Health Profile and Living Circumstances of Portuguese Centenarians from the Beira Interior Region. Curr Gerontol Geriatr Res. 2018;2018. doi:10.1155/2018/8450468

8. Kwon IS, Kim C-H, Ko HS, Cho S Il, Choi YH, Park SC. Risk factors of cardiovascular
930 disease in Korean exceptional longevity. J Korean Geriatr Soc. 2005;9: 251–265.

9. Franceschi C, Bonafè M. Centenarians as a model for healthy aging. Biochem Soc Trans.

2004; doi:10.1042/bst0310457

10. Heyn H, Li N, Ferreira HJ, Moran S, Pisano DG, Gomez A, et al. Distinct DNA methylomes of newborns and centenarians. Proc Natl Acad Sci. 2012; doi:10.1073/pnas.1120658109

11. Mondello C, Petropoulou C, Monti D, Gonos ES, Franceschi C, Nuzzo F. Telomere length in fibroblasts and blood cells from healthy centenarians. Exp Cell Res. 1999;248: 234–242. doi:10.1006/excr.1999.4398

12. Sebastiani P, Solovieff N, Puca A, Hartley SW, Melista E, Andersen S, et al. Genetic Signatures of Exceptional Longevity in Humans. Science (80-). 2010; doi:10.1126/science.1190532

13. Coale AJ, Kisker EE. Mortality crossovers: Reality or bad data? Popul Stud (NY). 1986; doi:10.1080/0032472031000142316

14. Fries JF. Aging, natural death, and the compression of morbidity. Bulletin of the World Health Organization. 2002. pp. 245–250. doi:10.1056/NEJM198007173030304

15. Vaupel JW, Manton KG, Stallard E. The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality. Demography. 1979;16: 439. doi:10.2307/2061224

16. Gavrilov LA, Gavrilova NS. Late-life mortality is underestimated because of data errors. PLoS Biol. 2019;17: e3000148. doi:https://doi.org/10.1371/journal.pbio.3000148

17. Preston SH, Elo IT, Stewart Q. Effects of age misreporting on mortality estimates at older ages. Popul Stud (NY). 1999; doi:10.1080/00324720308075

18. Gavrilova NS, Gavrilov LA. Mortality Trajectories at Extreme Old Ages: A Comparative Study of Different Data Sources on U.S. Old-Age Mortality. Living to 100 Monogr. 2014;2014. doi:10.1126/scisignal.2001449.Engineering

19. Newman SJ. Errors as a primary cause of late-life mortality deceleration and plateaus. PLoS Biol. 2018;16: e2006776. doi:https://doi.org/10.1371/journal.pbio.2006776

20. Newman SJ. Plane inclinations: A critique of hypothesis and model choice in Barbi et al. PLoS Biol. 2018;16: e3000048. doi:https://doi.org/10.1371/journal.pbio.3000048

21. Wachter KW. Hypothetical errors and plateaus: A response to Newman. PLoS Biol. 2018;16: e3000076. doi:https://doi.org/10.1371/journal.pbio.3000076

22. Farnsworth Riche M, Benton B, Schnelder PJ, Norton AJ. Population of the States and Counties of the United States: 1790 to 1990 [Internet]. Washington D.C.; 1996. Available: https://www.census.gov/population/www/censusdata/PopulationofStatesandCountiesofthe UnitedStates1790-1990.pdf

23. Gibson C, Jung K. Historical Census Statistics On Population Totals By Race, 1790 to 1990, and By Hispanic Origin, 1970 to 1990, For Large Cities And Other Urban Places In The United States. Washington D.C.; 2005.

24. Aldridge H, Theo BB, Tinson A, MacInnes T. London's Poverty Profile 2015. Tackling Poverty Inequal. 2015;

25. Ministry of Housing Communities & Local Government. English indices of deprivation 2019: research report [Internet]. 2019. Available: https://www.gov.uk/government/publications/english-indices-of-deprivation-2019-research-report

26. Pan American Health Organization. Country Report: French Guiana, Guadeloupe, and Martinique. In: Health in the Americas [Internet]. 2017 [cited 3 Feb 2020]. Available: https://www.paho.org/salud-en-las-americas-2017/?page_id=121

27. Barbi E, Lagona F, Marsili M, Vaupel JW, Wachter KW. The plateau of human mortality:

Demography of longevity pioneers. Science (80-). 2018;360: 1459–1461.

28. Timmer M, Baten J, Rijpma A, Smith C, van Zanden JL, Mira d'Ercole M. How Was Life?: Global Well-being since 1820. Paris: OECD publishing; 2014.

29. Willcox DC, Willcox BJ, He Q, Wang NC, Suzuki M. They really are that old: A validation study of centenarian prevalence in Okinawa. Journals Gerontol - Ser A Biol Sci Med Sci. 2008; doi:10.1093/gerona/63.4.338

30. Reuters. Italy's Dead Pensioners. New York Times. 1997: A00009.

31. Statistics Japan. e-Stat: Portal Site of Official Statistics of Japan [Internet]. 2020 [cited 30 Jan 2020]. Available: https://www.e-stat.go.jp/en

32. Director-General for Statistics and Information Policy, Ministry of Health L and W (organisation). Vital Statistics [Internet]. 2019 [cited 7 Feb 2019]. Available: https://www.mhlw.go.jp/english/database/db-hw/index.html

33. Poulain M. Exceptional longevity in Okinawa: A plea for in-depth validation. Demogr Res. 2011; doi:10.4054/DemRes.2011.25.7

34. Gavrilova NS, Gavrilov LA. Mortality Analysis of 1898–1902 Birth Cohort. Mortality Analysis of 1898–1902 Birth Cohort. Shaumburg, IL: Society of Actuaries.; 2018.

35. Koutantou A, Papachristou H. Greece pulls the plug on pensions for the dead. Reuters. 2012. Available: https://www.reuters.com/article/uk-greece-benefits/greece-pulls-the-plug-on-pensions-for-the-dead-idUKBRE83O14220120425

36. Teixeira L, Araújo L, Jopp D, Ribeiro O. Centenarians in Europe. Maturitas. 2017; doi:10.1016/j.maturitas.2017.08.005

37. Buettner D. The Blue Zones: lessons for living longer from those who've lived the longest. Washington, D.C.; 2010.

38. Motomura C. Global Agricultural Information Network Regional Report - Okinawa. 2014.

39. Tomuro K. Trends Observed in Poverty Rates, Working Poor Rates, Child Poverty Rates and Take-Up Rates of Public Assistance Across 47 Prefectures in Japan [Internet]. Yamagata University Faculty of Humanities Research Annual Report. Yamagata; 2016. Available: https://yamagata.repo.nii.ac.jp/?action=repository_uri&item_id=4067&file_id=17&file_no=1

40. Go M. Special feature: Thinking about modern society from Okinawa. Geppō shihōshoshi. 2004;June. Available: https://web.archive.org/web/20070825091736/http://www.shiho-shoshi.or.jp/web/publish/geppou/200406/2004_06_021.html

41. Cracolici MF, Uberti TE. Geographical distribution of crime in Italian provinces: A spatial econometric analysis. Jahrb fur Reg. 2009; doi:10.1007/s10037-008-0031-1

42. CIA. CIA World Factbook. World Factb 2013. 2017;

43. Arias E, Escobedo LA, Kennedy J, Fu C, Cisewski JA. U.S. small-area life expectancy estimates project : methodology and results summary [Internet]. (U.S.) NC for HS, editor. Hyattsville, MD; 2018. Available: https://stacks.cdc.gov/view/cdc/58853

44. Chetty R, Stepner M, Abraham S, Lin S, Scuderi B, Turner N, et al. The Association Between Income and Life Expectancy in the United States, 2001–2014. Jama. 2016;315: 1750–1766. doi:10.1001/jama.2016.4226

45. Poulain M, Herm A, Pes G. The blue zones: Areas of exceptional longevity around the world. Vienna Yearb Popul Res. 2013; doi:10.1553/populationyearbook2013s87

46. Elders.Stat database: Istituto Nazionale di Statistica. In: Istituto Nazionale di Statistica [Internet]. 2019 [cited 7 Feb 2019]. Available: http://dati-anziani.istat.it/

47. Gondo Y, Hirose N, Arai Y, Inagaki H, Masui Y, Yamamura K, et al. Functional status of centenarians in Tokyo, Japan: Developing better phenotypes of exceptional longevity. Journals Gerontol - Ser A Biol Sci Med Sci. 2006; doi:10.1093/gerona/61.3.305

48. Nakamura T, Tano J, Arai N, Jimbou K. Summary Report of Comprehensive Survey of Living Conditions 2016 [Internet]. 2017. Available: https://www.mhlw.go.jp/english/database/db-hss/dl/report_gaikyo_2016.pdf

49. Gellert C, Schöttker B, Brenner H. Smoking and All-Cause Mortality in Older People. Arch Intern Med. 2012; doi:10.1001/archinternmed.2012.1397

50. Taylor DH, Hasselblad V, Henley SJ, Thun MJ, Sloan FA. Benefits of smoking cessation for longevity. Am J Public Health. 2002; doi:10.2105/AJPH.92.6.990

51. Madans JH, Kleinman JC, Cox CS, Barbano HE, Feldman JJ, Cohen B, et al. 10 Years after NHANES I: Report of initial followup, 1982-1984. Public Health Rep. 1986;

52. Serdula MK, Koong SL, Williamson DF, Anda RF, Madans JH, Kleinman JC, et al. Alcohol intake and subsequent mortality: Findings from the NHANES I follow-up study. J Stud Alcohol. 1995; doi:10.15288/jsa.1995.56.233

53. Paganini-Hill A, Ross RK, Henderson BE. Prevalence of chronic disease and health practices in a retirement community. J Chronic Dis. 1986; doi:10.1016/0021-9681(86)90153-0

54. Paganini-Hill A, Kawas CH, Corrada MM. Type of alcohol consumed, changes in intake over time and mortality: The Leisure World Cohort Study. Age Ageing. 2007; doi:10.1093/ageing/afl184

55. Sebastiani P, Perls TT. The genetics of extreme longevity: Lessons from the new england centenarian study. Front Genet. 2012; doi:10.3389/fgene.2012.00277

56. Perls TT, K B, Freemen M, Alpert L, Silver M /H. Age Validation in the New England Centenarian Study. Validation of Exceptional Longevity. Odense; 1999.

57. Vaupel J, Vallin J, Meslé F, Robine J-M, Jdanov DA, Grigoriev O. The International Database on Longevity [Internet]. 2020 [cited 13 Feb 2020]. Available: https://www.supercentenarians.org/Home/Expand_IDL

58. United Nations. Coverage of Birth and Death Registration. In: United Nations Demographic Yearbook 2015: Quality of vital statistics obtained from civil registration [Internet]. 2017 [cited 25 Oct 2017]. Available: https://unstats.un.org/unsd/demographic/CRVS/CR_coverage.htm

59. Gondo Y, Hirose N, Yasumoto S, Arai Y, Saito Y. Age verification of the longest lived man in the world. Exp Gerontol. 2017; doi:10.1016/j.exger.2017.08.030

60. Takata Y, Ogawa G. Conscription system in Japan. New York, NY: Oxford University Press; 1921.

61. Gerontology Research Group. In: Table A: Verified Supercentenarians [Internet]. 2019 [cited 2 Jun 2019]. Available: http://www.grg.org/SC/SCListsTables.html

62. Levin J. Oldest American is 113-year-old Jersey girl. In: USA Today [Internet]. 2016 [cited 11 Feb 2020]. Available: https://www.usatoday.com/story/news/nation-now/2016/07/21/oldest-american-year-old-jersey-girl/87373424/

63. Maier H, Gampe J, Jeune B RJ and VJ. Supercentenarians. Demogr Res Monogr. 2010;7.

64. Smith D. RAF flypast for Europe's oldest man. In: The Guardian [Internet]. 2008 [cited 11 Feb 2020]. Available: https://www.theguardian.com/uk/2008/jun/01/military.firstworldwar

65. Christie J. The oldest living US veteran, age 108, credits drinking whiskey and smoking a

1070 dozen cigars every day for his long life. In: Daily Mail [Internet]. 2014 [cited 11 Feb 2020]. Available: https://www.dailymail.co.uk/news/article-2830888/The-oldest-living-veteran-smokes-drinks-whiskey-day-drives-does-yard-work-s-108-YEARS-OLD.html

66. Coles LS. Aging: The Reality: Demography of Human Supercentenarians. Journals Gerontol Ser A Biol Sci Med Sci. 2004;59: B579–B586. doi:10.1093/gerona/59.6.b579

1075 67. Levine ME, Crimmins EM. A Genetic Network Associated with Stress Resistance, Longevity, and Cancer in Humans. Journals Gerontol - Ser A Biol Sci Med Sci. 2016; doi:10.1093/gerona/glv141

68. Lv Y Bin, Liu S, Min Z-X, Gao X, Kraus VB, Mao C, et al. Associations of body mass index and waist circumference with 3-year all-cause mortality among the oldest old:

1080 evidence from a Chinese community-based prospective cohort study. J Am Med Dir Assoc. 2018;19: 672–678. doi:10.1016/j.jamda.2018.03.015

69. Oreopoulos A, Kalantar-Zadeh K, Sharma AM, Fonarow GC. The Obesity Paradox in the Elderly: Potential Mechanisms and Clinical Implications. Clin Geriatr Med. 2009;25: 643–659. doi:10.1016/j.cger.2009.07.005

1085 70. Japanese Ministry of Justice. About family register office work to affect location unknown elderly people [Internet]. 2010 [cited 18 Jun 2019]. Available: http://www.moj.go.jp/MINJI/minji04_00008.html.

71. Fleischer JG, Schulte R, Tsai HH, Tyagi S, Ibarra A, Shokhirev MN, et al. Predicting age from the transcriptome of human dermal fibroblasts. Genome Biol. 2018;

1090 doi:10.1186/s13059-018-1599-6

72. Malli RC, Aygun M, Ekenel HK. Apparent Age Estimation Using Ensemble of Deep Learning Models. IEEE Computer Society Conference on Computer Vision and Pattern

Recognition Workshops. 2016. doi:10.1109/CVPRW.2016.94

73. Ohtani S, Yamamoto T. Age estimation by amino acid racemization in human teeth. J Forensic Sci. 2010; doi:10.1111/j.1556-4029.2010.01472.x

74. Nielsen J, Hedeholm RB, Heinemeier J, Bushnell PG, Christiansen JS, Olsen J, et al. Eye lens radiocarbon reveals centuries of longevity in the Greenland shark (Somniosus microcephalus). Science (80-). 2016;353: 702–704. doi:10.1126/science.aaf1703

75. Hetzel AM. History and Organization of the Vital Statistics System. US Vital Statistics System: Major Activities and Developments, 1950-95. Centers for Disease Control and Prevention; National Center for Health Statistics; 1997. p. 75. Available: https://www.cdc.gov/nchs/data/misc/usvss.pdf

76. Bonneuil N. Table de mortalité France et par département et tables de migration nette par département, 1806-1906. Transformation of the French demographic landscape 1806 - 1906. Oxford: Clarendon Press; 1997.

77. Institut national de la statistique et des études économiques. Statistiques locales: Revenus-Pouvoir d'achat - comsommation Pauvreté [Internet]. 2016.

Acknowledgements

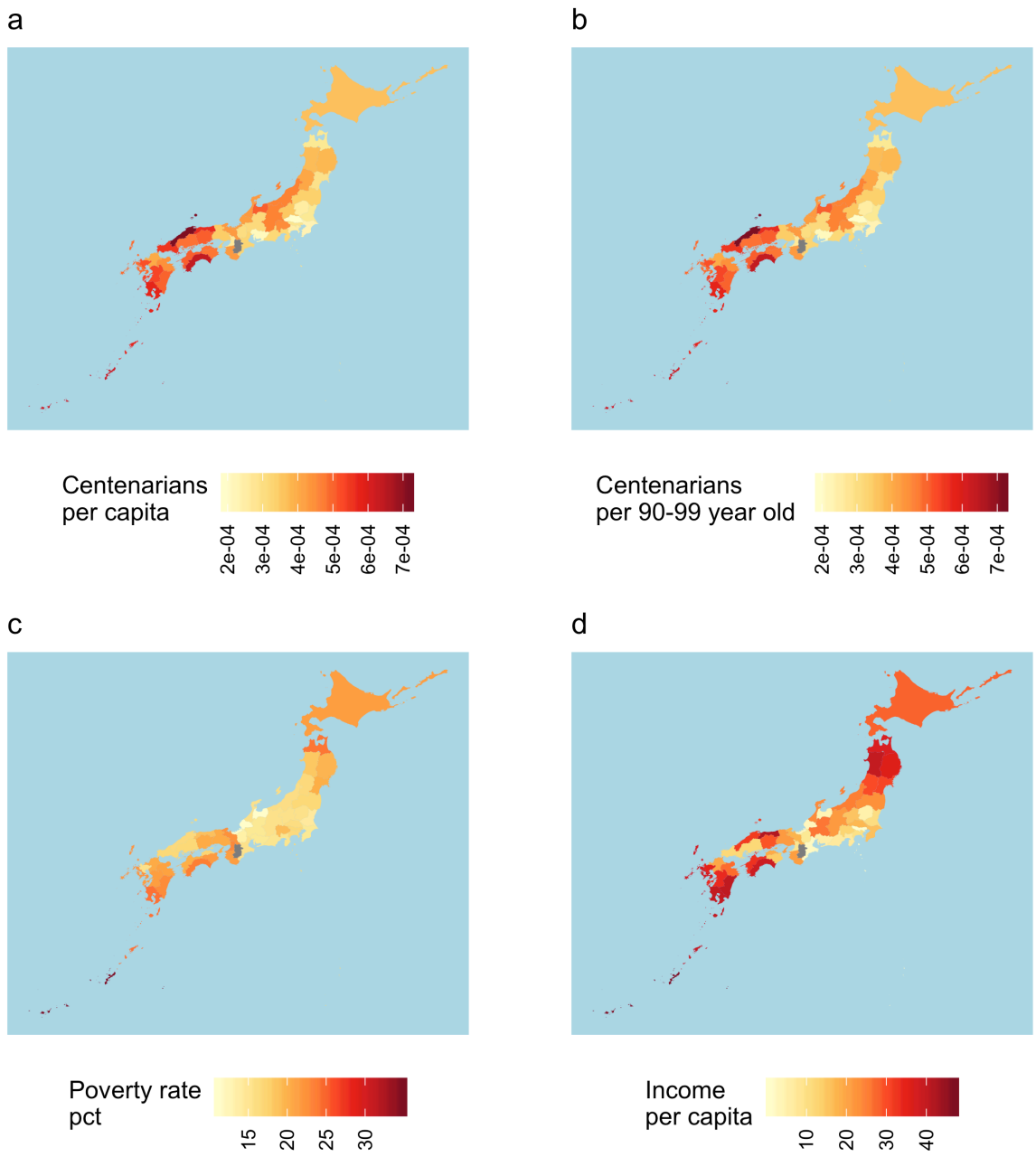
The author would like to acknowledge Zoe Campbell and Prof Heather Booth for providing much-needed feedback and editing advice, Chris Mulligan for providing inspiration to use US birth data, Sally Morell and Dr Jim Docherty for interesting leads, and a large section of the scientific community for providing commentary and support.

Author Contributions

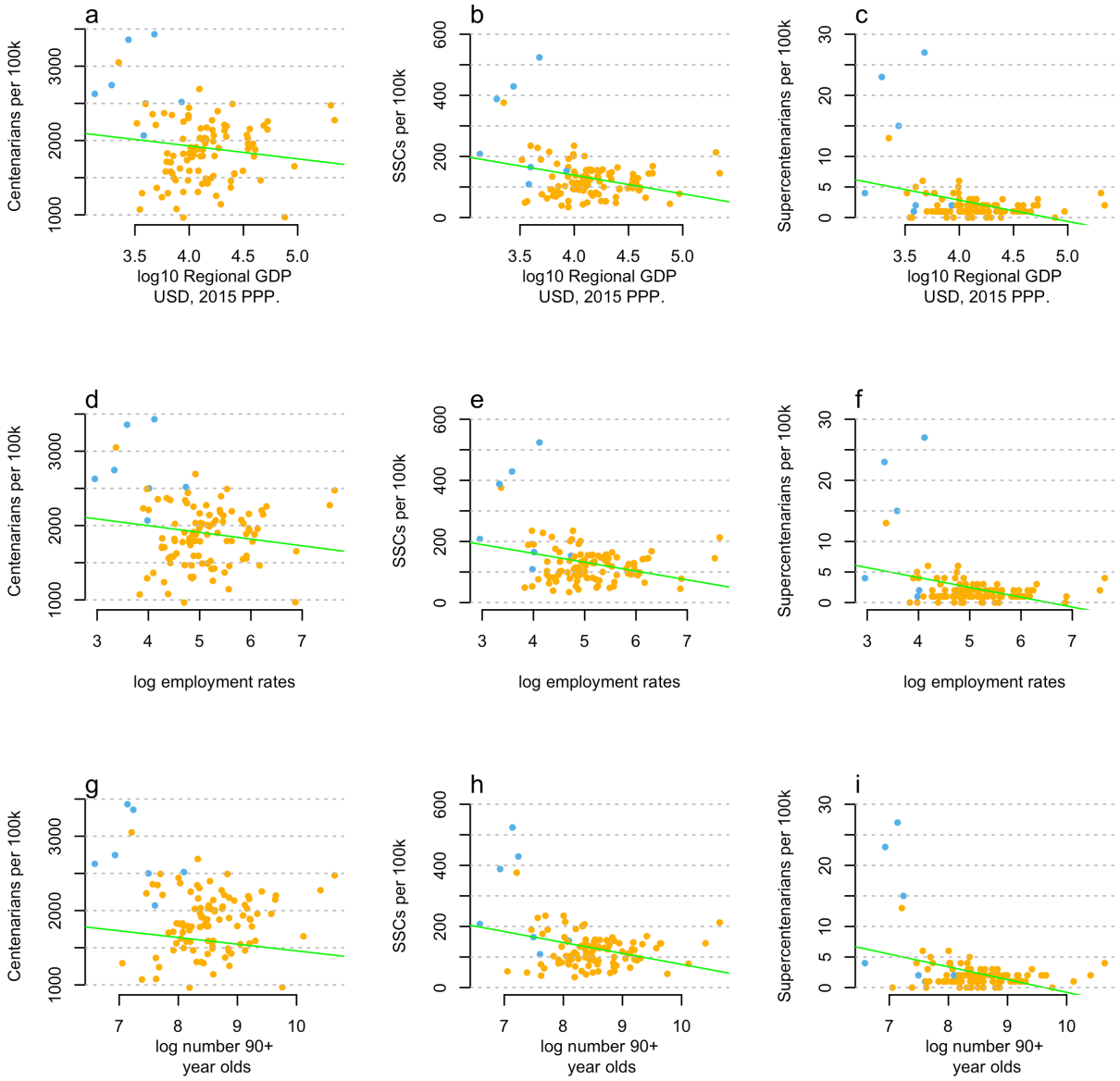
SJN conceived, designed, analyzed and wrote the study

Competing Interests

The author declares no competing interests.



1125 **Figure S1. Poverty and the distribution of Japanese Centenarians.** The 47 prefectures of Japan putatively contain over 48,000 centenarians, with generally higher concentration of centenarians per capita (a) and per 90-99 year-old person (b) in prefectures with high poverty rates (c) and lower-ranked prefectural incomes (d).



1130 **Figure S2. Italian provinces by GDP per capita and rates of extreme longevity.** According to figures from the Italian national statistics office and regional GDP data from the OECD, purchase-power parity adjusted GDP is negatively correlated with the frequency of (a) centenarians, (b) SSCs and (c) supercentenarians per capita across Italy: a pattern repeated in employment rates (d-f). Furthermore, the total number of 90+ year old people, shown here in log scale, is also negatively associated with the per

1135 capita number of centenarians (g), SSCs (h) and supercentenarians (i) across Italy. Linear mixed model regressions shown in green, Sardinian provinces shown in blue.

Figure S3. Historical life expectancy in the home region of birth for French supercentenarians. Supercentenarians are born in regions with an 84-day shorter life expectancy at birth, a non-significant reduction relative to the national average (one sample t-test NS; p = 0.52; N=143). Comparisons for supercentenarians born in overseas provinces and imperial holdings are unavailable, data show metropolitan France only.

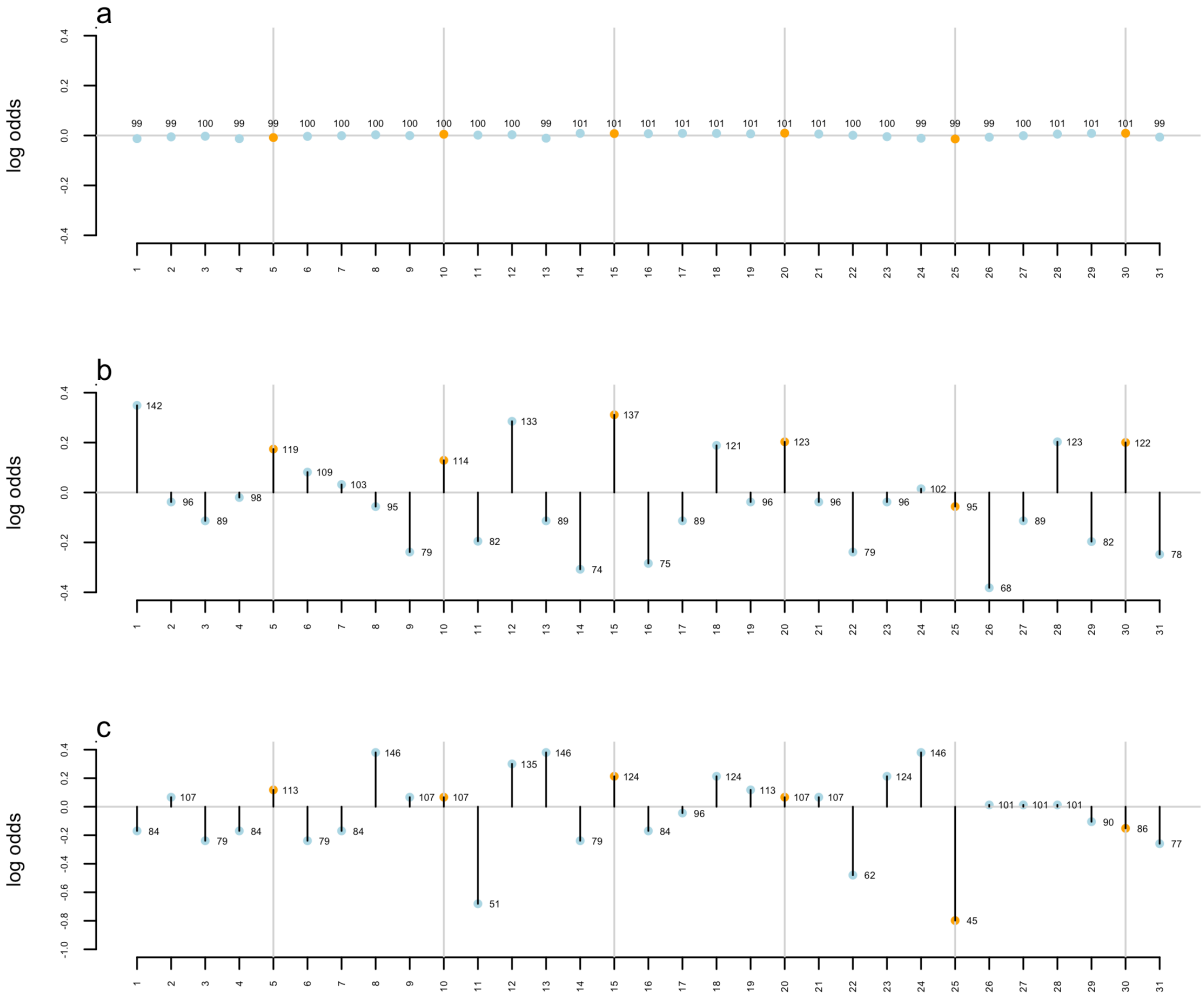


Figure S4. Age heaping of supercentenarian births. The distribution of modern birthdates (a), shown here by 70 million US birthdates observed from 1969-1988, display limited variation across days of the month. However, supercentenarian birth dates in the GRG (N = 1739) are 142% more likely to be born on the first day of the month and 118% more likely to be born on days that are multiples of five (orange points) compared with randomly distributed births (b). Age heaping on the first day of the month or in multiples of five is not as clear in the IDL data (c), possibly as a result of the removal of US birth days and months, or differences in cultural patterns (the 25th is heavily under-represented) and data quality. Points are labeled by the percentage of births over- or under-represented, relative to random sampling.

1155

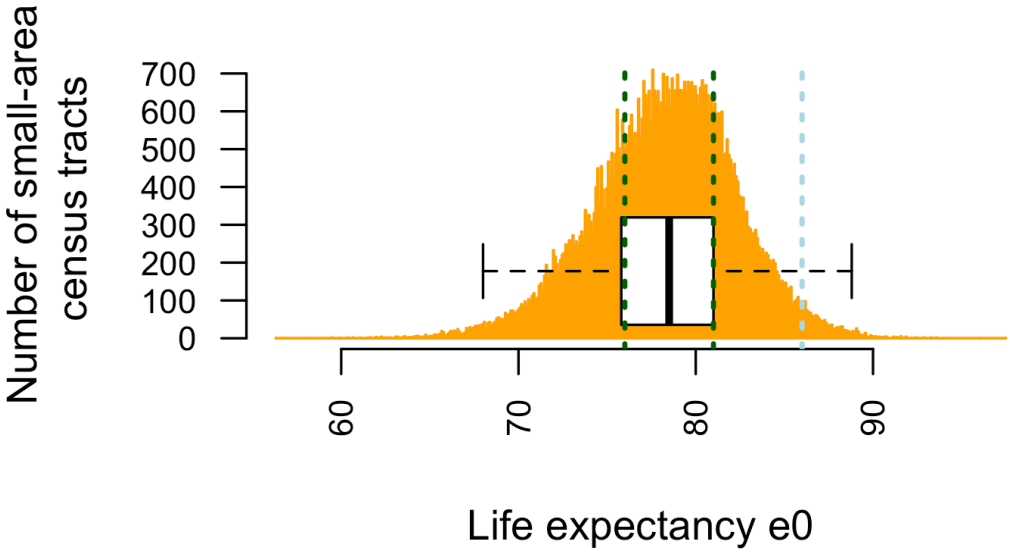


Figure S5. Distribution of life expectancy estimates in US small-area census tracts, including the Loma Linda 'blue zone'. When calculated independently by the Centers for Disease Control, the suburbs of Loma Linda range from the 27th to 75th percentiles of life

1160 expectancy at birth in the USA (green lines), relative to all other census tracts (orange). The absolute upper estimate, the female-only life expectancy calculated by 'blue zone' proponents (blue line), falls in the 98th percentile of life expectancy: high, but still behind 1401 longer-lived US census tracts and the nations of Japan, Singapore, Monaco, Spain, and South Korea.

1165

Table S1. Top 20 UK regions for the density of 105+ year old people, or SSCs, per capita.

			National Ranking, 1 = Worst, 131 = Best				
Region	**Total SSCs**	**SSCs per capita (rank)**	Income Deprivation Affecting Older People	90+ year-olds per capita (1=fewest)	Index of Multiple Deprivation	Crime rate	Health index
Tower Hamlets	15	1	1	1	26	23	52
Lewisham & Southwark	31	2	8	5	41	26	59
Darlington	5	3	62	85	46	14	32
Lambeth	15	4	5	3	49	18	56
Camden & City of London	11	5	55	18	90	119	110
Kensington & Chelsea and Hammersmith & Fulham	15	6	16	10	67	37	112
Isle of Wight	6	7	72	128	58	92	61
East Cumbria	11	8	120	121	94	130	82
Southampton	10	9	42	42	34	3	30
Portsmouth	8	10	44	65	33	10	48
Derby	9	11	48	64	40	75	33
North and West Norfolk	9	12	90	126	64	129	51
Telford and Wrekin	6	13	49	21	51	53	38
Manchester	18	14	3	6	4	1	3
Haringey & Islington	16	15	4	4	28	8	65
Leicester	11	16	7	17	17	25	27
Tyneside	27	17	29	62	25	39	16
Liverpool	15	18	6	16	2	12	2
Suffolk	23	19	101	120	86	91	97
East Kent	16	20	69	123	57	49	64

Table S2. Top 20 French regions for supercentenarians per capita (metropolitan and ranked overseas territories only).

				National Ranking, 1 = Worst, 101 = Best		
Region	**Super-centenarians (IDL)**	**Super-cent. per capita (rank)**	**Poverty rate ages 75+ (%)**	Poverty rate ages 75+	Poverty rate	GDP per capita, PPS adjusted
Creuse	3	1	17.0	4	16	4
Martinique	8	2	31.1	2	2	46
Guadeloupe	8	3	-	-	-	19
Cantal	2	4	13.9	10	60	21
Haute-Loire	3	5	11.1	26	77	10
Lozère	1	6	14.5	7	52	27
Lot	2	7	12.2	17	46	36
Indre	2	8	10.5	32	43	22
Yonne	3	9	7.0	75	44	39
Aude	3	10	14.4	8	5	11
Tarn	3	11	11.1	25	33	13
Paris	17	12	9.6	39	27	100
Jura	2	13	8.5	56	84	34
Maine-et-Loire	6	14	6.8	83	85	59
Aveyron	2	15	12.6	16	57	42
Hautes-Alpes	1	16	9.9	38	53	48
Orne	2	17	8.3	61	26	25
Ariège	1	18	12.8	15	14	6
Corse-du-Sud	1	19	15.9	5	19	75
Alpes-de-Haute-Provence	16	20	10.8	69	123	40

Table S3. Top 20 Italian provinces and regions for supercentenarians per capita.

				National Ranking, 1 = Worst 116 = Best			
Province	**Region**	**Super-centenarians per 100k (l_{110})**	**Rank l_{110}**	Survival to age 55	Unemployment	GDP per capita, PPS adjusted	90+ year-olds per capita (1=fewest)
Olbia-Tempio	Sardinia	27	1	7	14	41	8
Medio Campidano	Sardinia*	23	2	2	2	1	33
Carbonia-Iglesias	Sardinia	15	3	3	4	13	43
Isernia	Molise	13	4	4	3	34	102
Matera	Basilicata	6	5	61	15	17	23
Prato	Tuscany	6	6	114	40	80	59
L'Aquila	Abruzzo	5	7	65	39	49	82
Rieti	Lazio	5	8	8	7	31	80
Belluno	Veneto	4	9	45	29	96	91
Campobasso	Molise	4	10	38	22	35	76
Catanzaro	Calabria	4	11	59	52	26	31
Gorizia	Friuli-Venezia Giulia	4	12	34	10	64	107
Ogliastra	Sardinia*	4	13	1	1	23	66
Pistoia	Tuscany	4	14	87	37	59	84
Metropolitan City of Rome	Lazio	4	15	81	111	104	24
Vibo Valentia	Calabria	4	16	12	6	4	38
Brescia	Lombardy	3	17	90	106	94	25
Chieti	Abruzzo	3	18	64	64	53	71
Cremona	Lombardy	3	19	67	56	75	60
Fermo	Marche	3	20	35	-	-	-

* 'Blue zone' region

Table S4. Top 20 Japanese prefectures for centenarians per capita in 2015.

				National Ranking, 1 = Worst, 47 = Best		
Prefecture	**Centenarians (total)**	**Rank Centenarians per capita**	**Poverty rate (%)**	Poverty rate	Financial strength	Income per capita 2011
Shimane	515	1	16.7	30	1	13
Kochi	486	2	23.7	4	2	6
Okinawa*	872	3	34.8	1	5	1
Kagoshima	985	4	24.3	2	6	5
Tottori	334	5	18.9	21	3	2
Yamaguchi	806	6	16.9	27	23	36
Kumamoto	972	7	21.5	10	16	11
Saga	441	8	15.6	35	13	10
Toyama	554	9	11.2	47	25	44
Okayama	980	10	20.6	15	29	17
Ehime	720	11	20.2	17	20	12
Miyazaki	566	12	23.0	6	9	3
Nagasaki	700	13	22.2	8	7	7
Hiroshima	1395	14	16.9	26	35	35
Kagawa	482	15	17.2	24	27	28
Tokushima	377	16	21.8	9	8	33
Niigata	1105	17	16.0	33	19	21
Nagano	1000	18	15.5	36	26	25
Yamanashi	392	19	19.1	20	18	34
Oita	536	20	21.3	13	15	19

* 'Blue zone' region