BMJ Open

Inequalities in lifespan and mortality risk in the US, 2015–2019: a cross-sectional analysis of subpopulations by social determinants of health

Marie-Pier Bergeron-Boucher, Julia Callaway, Cosmo Strozza, Jim Oeppen

To cite: Bergeron-Boucher M-P, Callaway J, Strozza C, *et al*. Inequalities in lifespan and mortality risk in the US, 2015–2019: a cross-sectional analysis of subpopulations by social determinants of health. *BMJ Open* 2024;**14**:e079534. doi:10.1136/bmjopen-2023-079534

► Prepublication history and additional supplemental material for this article are available online. To view these files, please visit the journal online (https://doi.org/10.1136/bmjopen-2023-079534).

Received 04 September 2023
Accepted 26 April 2024

Check for updates

© Author(s) (or their employer(s)) 2024. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

Interdisciplinary Centre on Population Dynamics, Syddansk Universitet, Odense, Denmark

Correspondence to
Dr Marie-Pier Bergeron-Boucher; mpbergeron@sdu.dk

ABSTRACT

Objective To quantify inequalities in lifespan across multiple social determinants of health, how they act in tandem with one another, and to create a scoring system that can accurately identify subgroups of the population at high risk of mortality.

Design Comparison of life tables across 54 subpopulations defined by combinations of four social determinants of health: sex, marital status, education and race, using data from the Multiple Cause of Death dataset and the American Community Survey.

Setting United States, 2015–2019.

Main outcome measures We compared the partial life expectancies (PLEs) between age 30 and 90 years of all subpopulations. We also developed a scoring system to identify subgroups at high risk of mortality.

Results There is an 18.0-year difference between the subpopulations with the lowest and highest PLE. Differences in PLE between subpopulations are not significant in most pairwise comparisons. We visually illustrate how the PLE changes across social determinants of health. There is a complex interaction among social determinants of health, with no single determinant fully explaining the observed variation in lifespan. The proposed scoring system adds clarification to this interaction by yielding a single score that can be used to identify subgroups that might be at high risk of mortality. A similar scoring system by cause of death was also created to identify which subgroups could be considered at high risk of mortality from specific causes. Even if subgroups have similar mortality levels, they are often subject to different cause-specific mortality risks.

Conclusions Having one characteristic associated with higher mortality is often not sufficient to be considered at high risk of mortality, but the risk increases with the number of such characteristics. Reducing inequalities is vital for societies, and better identifying individuals and subgroups at high risk of mortality is necessary for public health policy.

STRENGHTS AND LIMITATIONS OF THIS STUDY

- ⇒ Quantifies the intersection of sociodemographic factors on risk of mortality.
- ⇒ Develops a scoring system that can efficiently identify subgroups that are at high risk of mortality.
- ⇒ Summarises mortality levels across social determinants of health.
- ⇒ Based on cross-sectional data.
- ⇒ Uses a limited number of sociodemographic variables in the analysis.

INTRODUCTION

Who is at high risk of mortality? This question has been the focus of many public health, actuarial and social policies. Identifying at-risk individuals can nevertheless be complex, as multiple factors influence health and mortality. Although individual risks and genetic factors explain part of the differences in health and mortality, a large and increasing body of evidence reveals the role of social factors in shaping health.[1] These social determinants of health are non-medical factors that influence health, comprising "conditions in which people are born, grow, live, work and age".[2] They include factors such as income, education, race, place of residence, employment status and social support networks. In some settings, social determinants have been found to have a larger impact on health than medical care.[1] This is seen, for instance, in Galea *et al* (2011), who found the number of deaths attributable to low education in the United States (US) in 2000 to be similar to that from myocardial infarctions.[3] Individuals with lower education and income levels have been found to have lower life expectancies than those with higher levels,[4 5] with these differences increasing over time.[6–8] Likewise, many studies reveal a social gradient in health and mortality across other factors in addition to education and income, including sex, race, marital status and place of residence.[1] For instance, men have lower life expectancies than women,[9 10] those unmarried lower than those married,[11] and Blacks lower than

BMJ Open: first published as 10.1136/bmjopen-2023-079534 on 5 August 2024. Downloaded from http://bmjopen.bmj.com/ on September 9, 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Whites.[12 13] Analysing race in public health research represents the interplay of structural factors that leads to racial and ethnic inequalities in health, health outcomes and mortality. These factors include housing, education, employment, environment, earnings, benefits, credit, media, healthcare and criminal justice, and are the basis of structural racism in the US.[14]

The social gradient in health and mortality is often quantified by comparing summary measures, for example, life expectancy and standardised mortality rates. On average, individuals of higher socioeconomic status (SES) have higher life expectancy and lower mortality than individuals from lower SES groups. These measures quantify between-group inequalities, whereas within-group inequalities, often referred to as lifespan inequalities, explain the variation of lifespan around these summary measures.[15] Lifespan inequality captures how equal the lifespans are within a population, with lifespan distribution often having a large variation. Research on lifespan inequalities reveals overlaps in lifespan distributions of two or more populations, meaning some individuals from a population with a low life expectancy live longer than some individuals from a population with a high life expectancy. Studies that measure overlaps in distributions are limited[16–18]; however, a small but emerging body of research shows some of the characteristics of these overlapping lifespan distributions.[17 18] These overlaps occur because there is lifespan inequality within each compared population, which reflects remaining variation in individual, health and social characteristics. For example, women, on average, live longer than men, but not all women and men have the same education level or marital status, and some men end up having long lifespans while some women end up having short lifespans. So, is being male sufficient to be considered high risk for mortality? The answer is no. Research reveals that the male survival disadvantage is driven by some subpopulations of men with particularly high mortality.[19]

Public health policies are often designed to benefit a target group impacted by a disease or a health condition. In public health, defining a target group is "an important condition for formulating realistic objectives for reaching these objectives as well as for reaching the group itself".[20] When it comes to identifying groups at high risk of mortality, using only one social determinant of health does not provide well-defined groups, as large lifespan inequalities can remain within these groups. To understand these inequalities, how they act in tandem with one another, and to identify individuals at high risk of mortality more accurately, we investigate lifespan differences for subpopulations based on multiple social determinants of health – sex, race, marital status and education level – and propose a scoring system to assess overall and cause-specific mortality risk more accurately. The scoring system proposed in this article can reflect the intersectional nature of mortality risk. Intersectionality is a theoretical framework whereby multiple socioeconomic factors intersect at the individual level to reflect inequalities at the population level.[21] In the context of public health research, intersectionality can be analysed to understand inequalities in health outcomes and mortality. The scoring system allows one to more easily assess how multiple factors intersect or interact to lead to inequalities in lifespan.

METHODS

Data

Death counts were extracted from the Multiple Cause of Death dataset from the National Vital Statistics System of the National Center for Health Statistics.[22] Deaths were provided by single-year age groups, sex, marital status, education and race. Population counts were extracted from the American Community Survey (ACS) from the United States Census Bureau.[23] The ACS has data with similar variables and single-year age groups until age 99 years. We used pooled data from 2015 to 2019. We did not include data from 2020 and onwards to avoid the effect of the COVID-19 pandemic.

We analysed all combinations of sex, marital status, education and race, categorised as:

- Sex – female (F) or male (M).
- Marital status – married (Ma), previously married (Pm) (divorced and widowed) or never married (Nm).
- Education – high school diploma or less (Hd), some college or associate degree (Sa) or university degree (Ud).
- Race – Black (B), Hispanic White (H) or non-Hispanic White (W).

In total, 54 US subpopulations were analysed.

The open-ended age group within the ACS varies across states and survey year. We found that the analysis of old-age mortality can only be done for an open age group of 90+ years. This issue, combined with the low population counts at older ages, and the "salmon bias", whereby older Hispanics leave the US when their health begins to deteriorate, leaving behind a healthier Hispanic population than there would have been otherwise, leads to estimation problems of old-age mortality for many of the subpopulations. Non-robust estimates of mortality and life expectancy at age 90+ years were found. To avoid potential bias we limit the analysis to age 30 to 89 years.

We also analysed cause-specific mortality trends. The Multiple Cause of Death dataset offers such data for the US. We looked at 12 underlying causes of death, representing the 11 leading causes of death in the US in 2019 and homicide.[24] We included homicide as this cause is an important contributor to lifespan inequalities. Cancers were broken down into four groups: lung cancer, breast cancer, prostate cancer and other cancers. The causes of death and their codes from the 10th Revision of the International Classification of Diseases (ICD-10) are shown in the online supplemental material.

BMJ Open: first published as 10.1136/bmjopen-2023-079534 on 5 August 2024. Downloaded from http://bmjopen.bmj.com/ on September 9, 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Life tables

Life tables were calculated for all possible combinations of determinants, totaling 54 subpopulations (two sexes × three marital statuses × three education levels × three races). We used the period (cross-sectional) life table, which is used to calculate the life expectancy of a hypothetical cohort of individuals, as if individuals born in a given year experienced the age-specific mortality rates observed that year over their lifetime. Period life expectancy is one of the most used measures of population health.[25] However, due to data problems at older ages, we limit the analysis to mortality and partial life expectancy (PLE) between age 30 and 90 years. PLE measures the expected number of years lived between specific ages, here between age 30 and 90 years for a maximum of 60 years of life. Confidence intervals for the life tables were calculated using the Chiang method.[26]

Decomposition

To understand the differences in PLE between subpopulations, we applied the Arriaga decomposition by cause of death.[27 28] The method decomposes the contribution of specific causes of death to the absolute difference in life expectancy. We adapted the method to decompose PLE instead of life expectancy.

Scoring systems

Considering several social characteristics increases the accuracy, but also the complexity, of the analysis required to identify individuals at high risk of mortality. Scoring systems are an efficient solution to this problem because they summarise data from multiple variables to yield a single score. For example, the Apgar score uses five criteria from an infant at 1 minute of age to establish the need for breathing intervention.[29 30] The Framingham Risk Score uses a combination of behavioural, clinical and demographic criteria to yield a score that estimates the 10-year cardiovascular risk of an individual.[31 32] Several scoring systems were developed to predict COVID-19 mortality risk during the pandemic.[33 34] From a public health perspective, these scoring systems are useful tools that can rapidly identify groups at higher risk of mortality or other health outcomes based on several variables. Many methods have been suggested to establish a scoring system, most of which rely on regression analysis. The Framingham Risk Score is based on the beta coefficients of an age-adjusted Cox proportional hazards regression.[31 32] Scoring systems to predict COVID-19 mortality have been established based on a logistic regression model or stepwise regression.[33 34]

In this article, we suggest a lifespan scoring system considering social determinants of health, identifying the risk for an individual to be subject to high mortality rates. The system is based on a Cox proportional hazard model applied to the age-specific death rates of the subpopulations. The scores are the standardised (beta) coefficients of the Cox proportional hazard model multiplied by −10 and rounded. We use the population with the closest age-specific death rates to the national population as reference in the Cox proportional hazard model (Black married men with some college or an associate degree). We chose to multiply the coefficient by −10 to provide a single unit score. A negative score indicates that the population has worse survival (higher mortality) than the average and a positive score indicates better survival (low mortality). As the method is based on the Cox proportional hazard model, the coefficients have the same interpretation: they are the logged hazard ratios (HRs), with $\exp(\beta/10)$ expressing by how many times the studied population has a higher or lower mortality than the average. The system was compared with another system based on a linear regression applied to the PLE (see online supplemental material) and very similar scores were found, despite the Cox proportional hazard model capturing the logged HR, and the latter model capturing differences in PLE (coefficient also multiplied by 10). In this case, the hazard scale can then easily be converted into differences in years of life, a more easily interpretable metric. The scoring system is based on cross-sectional data, representing a mortality risk for the life tables' hypothetical cohorts.

The variable scores are additive. Although we decided not to add the complexity of interaction effects to the scoring system (but a model including interaction can be found in the online supplemental material), the hypothesis that interactions of sex, race and educational attainment can be multiplicative rather than additive is confirmed by Mehta and Preston[35] in a relative risk model for US mortality. They show that such interactions are testable, but confirm that they are difficult to interpret.

Patient and public involvement

No patients or members of the public were involved in the study.

RESULTS

Our analysis is based on pooled data over 5 years (2015–2019) that we divide into 54 subpopulations. We observe population sizes ranging from a minimum of 1.02 million (obtained applying sample weights to 8634 individuals) and 11 014 deaths for Hispanic men with a university degree who were previously married, to a maximum of 8.34 million (obtained applying sample weights to 0.96 million individuals) and about 0.7 million deaths for White men with a university degree who are married. We include information on population size and deaths for all the 54 subpopulations in the supplementary material. Figure 1 shows the PLE between age 30 and 90 years for the 54 subpopulations. Two main results emerge from this figure. First, there are very large differences in PLE between the subpopulations. There is an 18.0 year difference between the minimum and maximum PLE between age 30 and 90 years across the subpopulations. White never married men with a high school diploma or less have the lowest PLE, with 37.1 years. By contrast, White

BMJ Open: first published as 10.1136/bmjopen-2023-079534 on 5 August 2024. Downloaded from http://bmjopen.bmj.com/ on September 9, 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

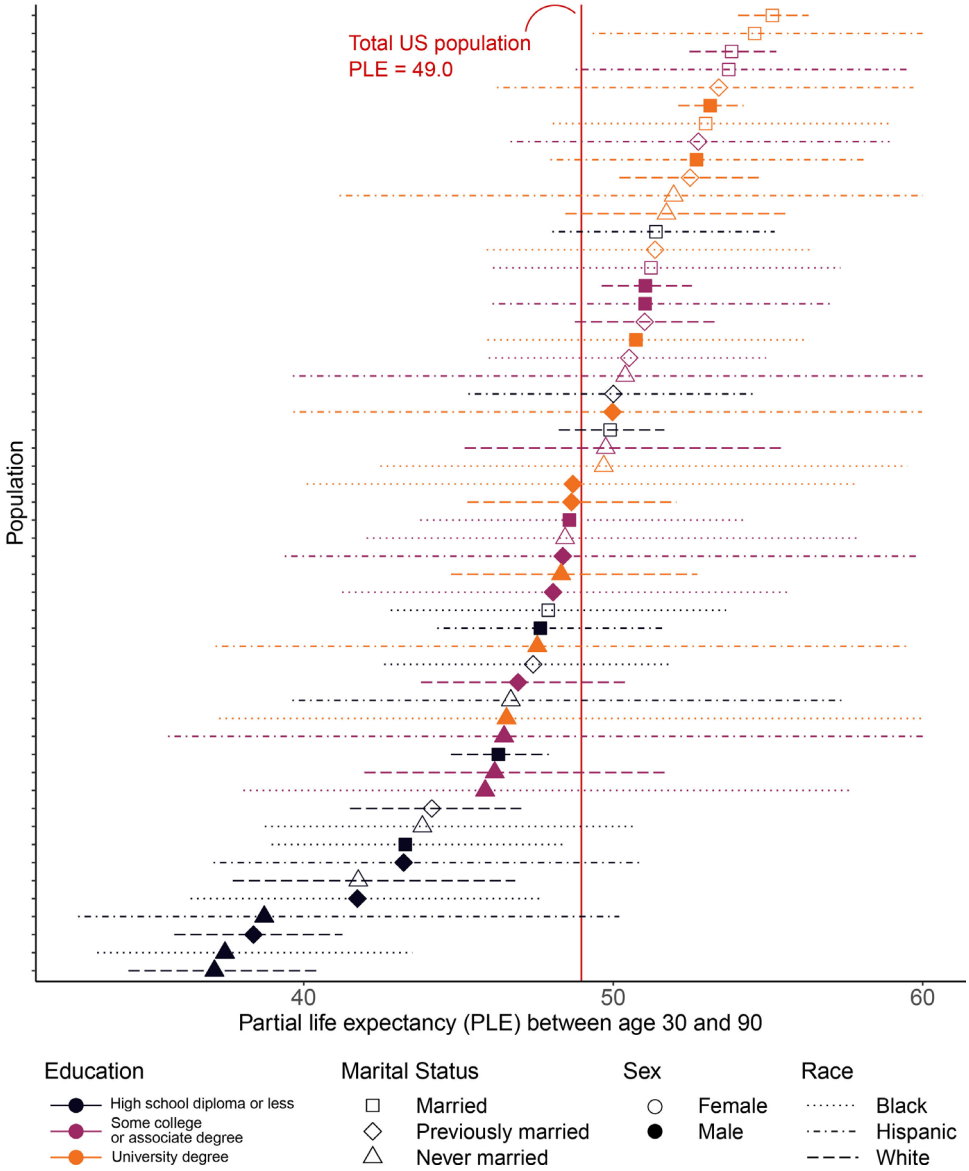

Figure 1 Partial life expectancy between age 30 and 90 years with 95% confidence interval by social determinants of health, United States 2015–2019. Sources: Multiple Cause of Death Data (MCDD),[22] American Community Survey (ACS)[23] and authors' own calculation.

married women with a university degree have the highest PLE, with 55.1 years. Due to the relatively small sample sizes in the ACS, the confidence intervals of the subpopulations are large. The difference in PLE between the lowest and highest populations is significantly different. However, no significant differences are found in most pairwise comparisons.

Second, figure 1 shows that the advantage of groups with high PLEs is not obvious once multiple determinants are considered. No one determinant brings a clear advantage to all individuals equally. For example, some subpopulations of women have lower PLEs than some subpopulations of men. White married men with a university degree have a PLE of 53.2 years, which is higher than 81% of all the women subpopulations (22 of 27 female groups). Similarly, some subpopulations with low education outlive some with high education. Married Hispanic women with a high school diploma or less have a PLE of 51.4 years, which is higher than 44% of the subpopulations characterised by a university degree (8 of 18 groups with a university degree).

There are many factors that influence health and mortality, and different combinations of social determinants lead to different PLEs. Figure 2 illustrates the complexity of the contribution of social determinants of health to the length of life. The vertical axis represents the PLE for the subpopulations. The centre of the network is the PLE for the total US population (red circle) with a value of 49.0 years. The first partition is the most important variable that explains the differences in lifespan, here being education level. Each coloured circle represents the PLE for a given education level. The next partition is the second most important variable (marital status) and so on. For example, having a

BMJ Open: first published as 10.1136/bmjopen-2023-079534 on 5 August 2024. Downloaded from http://bmjopen.bmj.com/ on September 9, 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

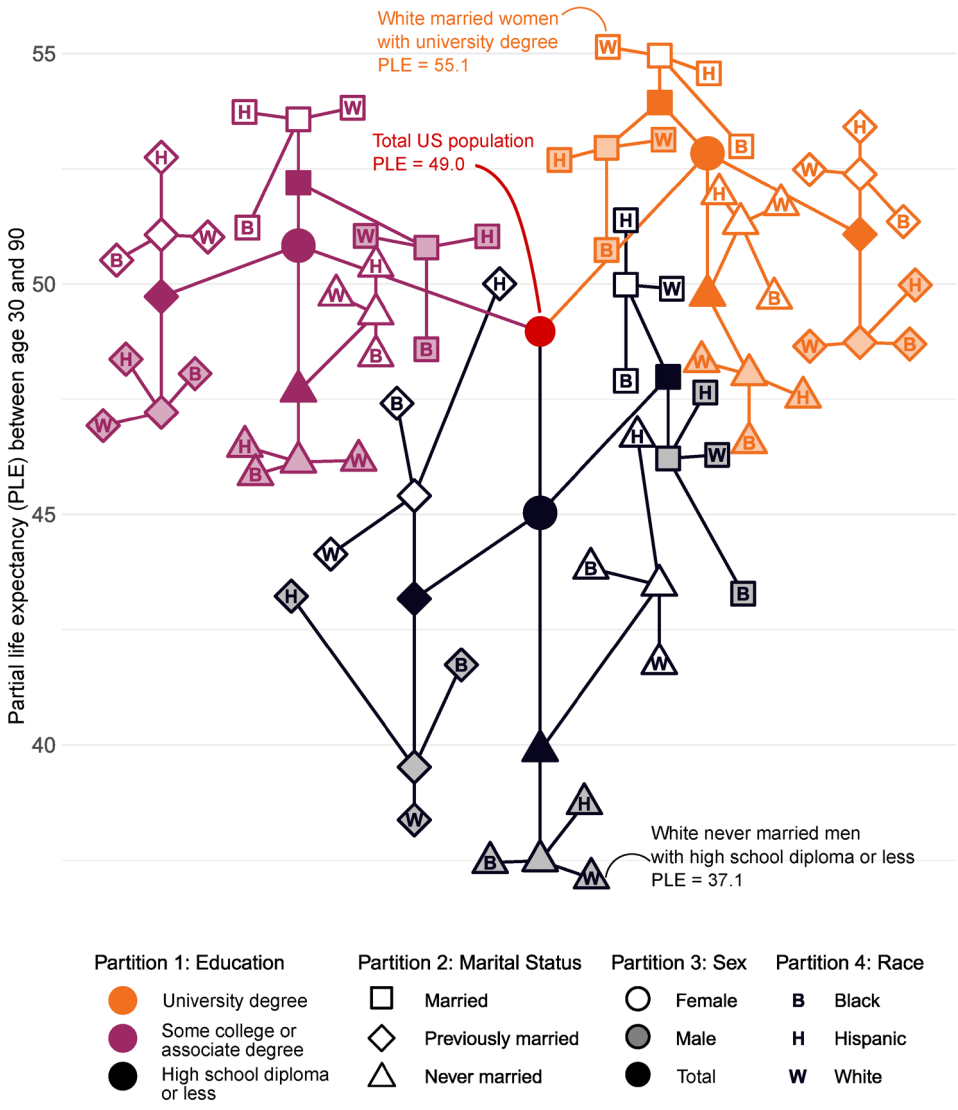

Figure 2 Tree of partial life expectancy between age 30 and 90 years across social determinants of health, United States. 2015–2019. Sources: Multiple Cause of Death Data (MCDD),[22] American Community Survey (ACS)[23] and authors' own calculation.

high school diploma or less (in black) reduces the PLE to 45.0 years, but following the path through married (squares), female (no fill) and Hispanic (H) shows that this combination has a PLE of 51.4 years. The figure illustrates that a characteristic that negatively impacts PLE can be offset by a characteristic that positively impacts PLE and vice versa. For example, having a high school diploma or less reduces PLE by 3.9 years compared with the total. However, being married and female increases PLE by 4.9 years, bringing the life expectancy of married women with a high school diploma or less to above the national PLE. To give another example, having a university degree increases PLE by 3.9 years compared with the total, but being never married and male decreases the PLE of highly educated individuals by 4.8 years, bringing the PLE of never married men with a university degree below the national level. What brings an individual to the top (or the bottom) of the scale is not having one characteristic that impacts PLE positively (or negatively) but having more than one. The subpopulations with the lowest PLEs are those with multiple social determinants that are associated with worse health.

At what point does an individual become at high risk of mortality and short lifespan? In table 1, we present a scoring system for identifying groups at high risk of mortality across social determinants of health. Each variable score is additive. For example, being female (score of 4), married (0), White (1) and having a high school diploma or less (−5) yield a total score of 0. Based on this system, the total scores can take values between −10 and 8. The definition of high mortality can vary depending on the criteria. For example, if high mortality were defined as being higher than the average, then subpopulations with a total score of less than 0 would be considered at

BMJ Open: first published as 10.1136/bmjopen-2023-079534 on 5 August 2024. Downloaded from http://bmjopen.bmj.com/ on September 9, 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Table 1 Additive lifespan scoring system across social determinants of health, United States 2015–2019

Social determinant	−5	−4	−3	−2	−1	0	1	2	3	4
Sex						M				F
Marital	Nm		Pm			Ma				
Education	Hd					Sa		Ud		
Race						B	W	H		

Note: Sex: female (F), male (M). Marital status: married (Ma), previously married (Pm), never married (Nm). Education: high school diploma or less (Hd), some college or associate degree (Sa), university degree (Ud). Race: Black (B), Hispanic-White (H), non-Hispanic White (W).
Sources: Multiple Cause of Death Data (MCDD),[22] American Community Survey (ACS)[23] and authors' own calculation.

high risk. The score of 0 is obtained by White married women with a high school degree or less, White never married women with some college or an associate degree, Black married men with some college or an associate degree, and White previously married men with a university degree. Conversely, if high mortality were defined as having death rates one standard deviation (SD) above the mean, then a score of −5 would indicate a high mortality risk. About 50% of the subpopulations have a score of 0 or higher and 19% of −5 or lower. However, this does not mean that individuals in subpopulations that have a low score will not survive to older ages, as a large variation within groups remains (see online supplemental material). However, it does indicate that individuals with these characteristics might be subject to higher mortality risk, and thus might require more medical or public health intervention.

Figure 3 shows the contribution, in years, of various causes of death to the difference in PLE among the subpopulations and the national level. Subpopulations are ordered according to their PLE, increasing from bottom to top. It shows how many years of life each cause contributes negatively or positively to each subpopulation, relative to the national level. Overall, mortality differences from heart disease, unintentional injuries (U. injuries) and other causes explain most of the difference in PLE between the subpopulations. Subpopulations with low PLEs are those with high mortality rates from these three causes of death. The figure also shows that for the least advantaged subpopulations, most causes have a negative contribution to the PLE, and for the most advantaged groups, most causes have a positive contribution. However, the picture is not as clear for subpopulations in the centre of the ranking. Some causes have positive contributions and others have negative contributions for different subpopulations.

Figure 4 illustrates a scoring system for mortality by cause of death, using a similar methodology to the scoring system introduced in table 1. This system can indicate if some subpopulations are at higher risk of mortality from specific diseases, relative to the weighted average cause-specific mortality rate. The range of the scores varies between causes. We found a positive correlation between the range of the scores by cause and the SD of the death rates across subpopulations, with a larger SD coming with a larger range of scores (see online supplemental material). For example, for suicide, the scores can take values between −4 and 12 (range of 16), and the mean SD for the death rates across ages is 1.12. In comparison, the scores for breast cancer can vary between −6 and 3 (range of 9), with a mean SD of 0.39. The causes of death in figure 4 are ordered by SD, from the cause with the largest SD across subpopulations to the smallest.

The range of the scores for each social determinant is also representative of its importance in explaining the mortality differences among subpopulations. For instance, education explains more of the variation across subpopulations for chronic lower respiratory diseases (CLRD) (range of 7) than it does for suicide (range of 1). Similarly, there were more differences between men and women for suicides than for CLRD.

Figure 4 also shows that while being married always lowers mortality compared with not being married, and that higher education always lowers mortality compared with lower education levels, the gradient is not as obvious for race. Some causes negatively impact more Whites, such as suicide, unintentional injuries, CLRD and lung cancer, while others negatively impact more Hispanics, such as liver diseases. However, most causes tend to negatively impact Blacks across the board. Women generally have a survival advantage over men across all causes, except for other cancers and Alzheimer's disease.

Different social determinants of health have different impacts on each cause of death. As a result, the causes of death contributing to the mortality disadvantage vary across the subpopulations, even if the overall score is the same (see online supplemental material). For example, White previously married women with a high school degree or less are at high risk of mortality from lung cancer, other cancers, CLRD, Alzheimer's disease, kidney diseases, and influenza and pneumonia. Black never married men with a university degree are at higher risk of mortality from heart disease, prostate cancer, cerebrovascular diseases, diabetes, kidney diseases and homicide. Both subpopulations have an overall mortality score of −3.

DISCUSSION

There is a complex interaction between social and individual determinants of health, with no one determinant explaining the full observed variation in lifespan. Having one characteristic that is associated with higher mortality is often not a sufficient criterion to be considered at high risk of mortality, but the risk does increase with the number of such characteristics. In addition, not all analysed social determinants of health have the same degree of influence on lifespan and mortality. For example,

BMJ Open: first published as 10.1136/bmjopen-2023-079534 on 5 August 2024. Downloaded from http://bmjopen.bmj.com/ on September 9, 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

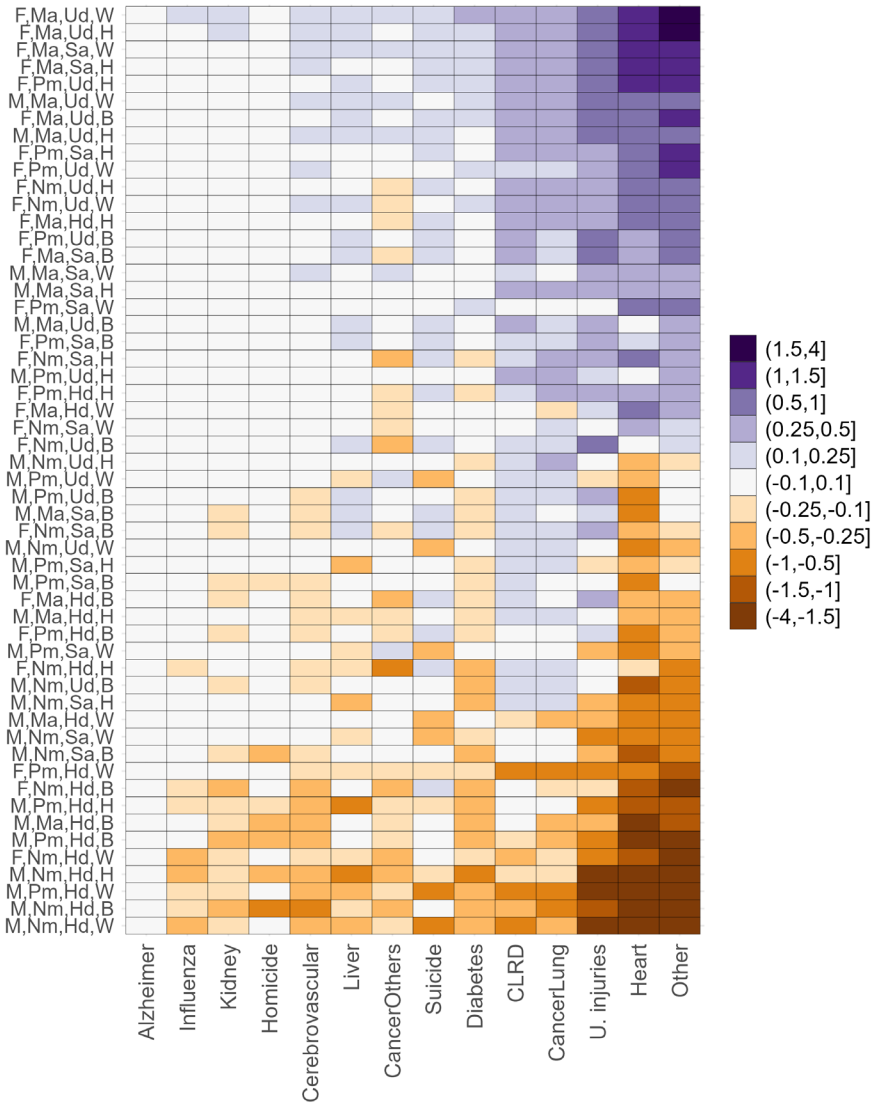

Figure 3 Contribution in years from various causes of death to the difference in partial life expectancy between age 30 and 90 years for 54 subpopulations characterised by various social determinants of health and the national value, United States 2015–2019. Note: Results for breast and prostate cancer are not shown here. As they are sex-specific types of cancer, and the comparison is with the total US population, women are automatically at high risk of breast cancer and men of prostate cancer. Sex: female (F), male (M). Marital status: married (Ma), previously married (Pm), never married (Nm). Education: high school diploma or less (Hd), some college or associate degree (Sa), university degree (Ud). Race: Black (B), Hispanic-White (H), non-Hispanic White (W). Sources: Multiple Cause of Death Data (MCDD),[22] American Community Survey (ACS)[23] and authors' own calculation.

we found that education often has a greater impact on lifespan differentials than race. Multiple factors influence health in various manners, making it difficult to identify at-risk individuals. We introduced a scoring system to help identify individuals at high risk of mortality based on their social determinants of health in a simple manner.

The use of PLE and its graphical representation provide some advantages: (1) it converts a hazard scale to years of life (we showed that models based on hazard and PLE provide similar scores) and (2) it is easier to understand, as one does not have to add together and exponentiate the parameters to understand how factors in the model were combined to exacerbate or offset risks.

Limitations

The scoring system could oversimplify the underlying factors at play. We only considered four characteristics – education, race, sex and marital status – due to the limited number of variables available in the selected database. Including more variables, such as income, place of residence, environmental factors, access to healthcare, or information on health-related behaviours (eg, smoking) could make the scoring system more precise and efficient. It is important to note that despite decomposing the US population into 54 subpopulations, large within-population inequalities remain (see online supplemental material). Many additional factors contribute to

BMJ Open: first published as 10.1136/bmjopen-2023-079534 on 5 August 2024. Downloaded from http://bmjopen.bmj.com/ on September 9, 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

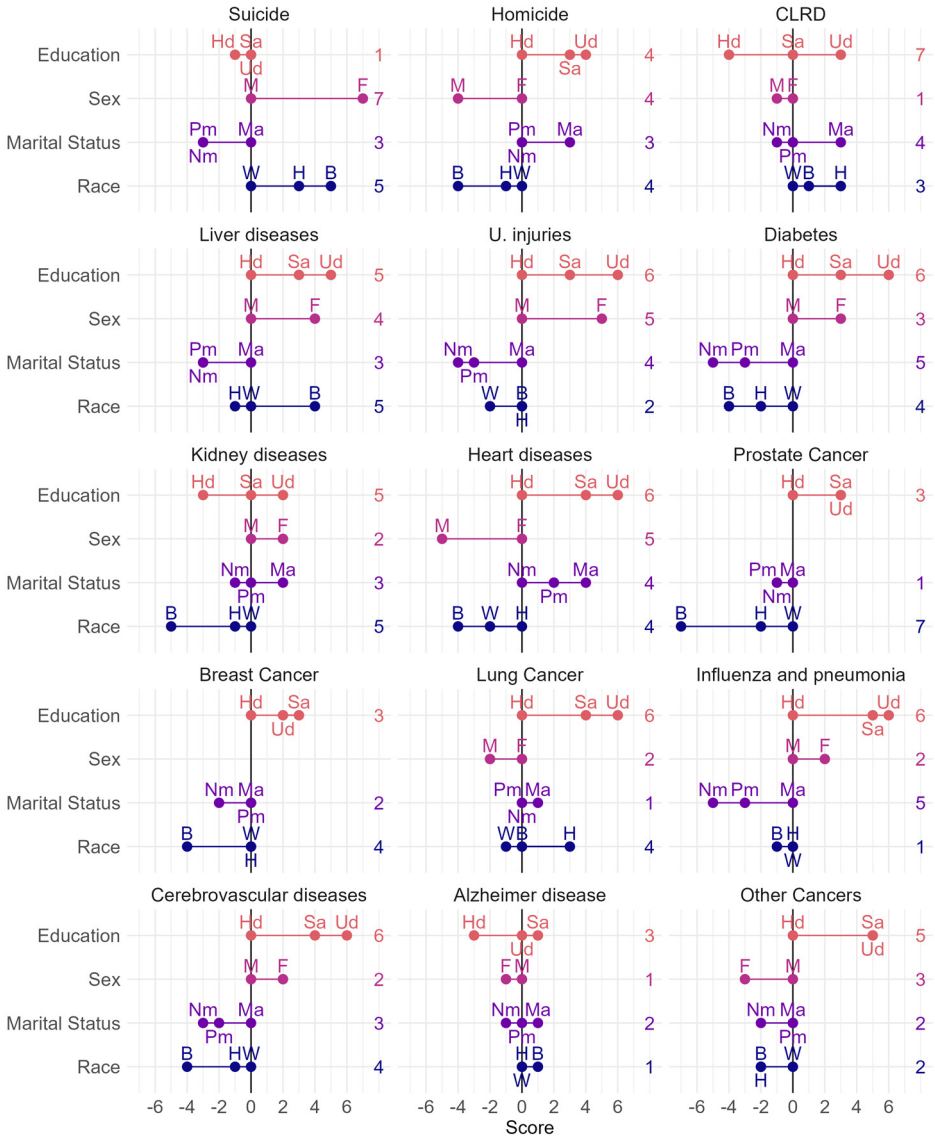

Figure 4 Scoring system across social determinants of heath by cause of death, United States 2015–2019. The variable scores are additive. The range of the score of each variable is written at the right of each panel. Note: For breast cancer, the analysis is only performed for women; and only for men for prostate cancer. Sex: female (F), male (M). Marital status: married (Ma), previously married (Pm), never married (Nm). Education: high school diploma or less (Hd), some college or associate degree (Sa), university degree (Ud). Race: Black (B), Hispanic-White (H), non-Hispanic White (W). CLRD, chronic lower respiratory diseases; U. injuries, unintentional injuries. Sources: Multiple Cause of Death Data (MCDD),[22] American Community Survey (ACS)[23] and authors' own calculation.

inequalities in lifespan. As a result, the lifespan distributions of all subpopulations overlap. The probability that White never married men with a high school diploma or less will outlive a random person (using the national lifespan distribution) between age 30 and 90 years in the US is 29%. At the other end of the spectrum, White married women with a university degree have a 63% probability of outliving a random person in the US. The probability that an individual from the former subpopulation will outlive an individual in the latter is 19%. While the limited number of characteristics also limits the explained variance of lifespan, a smaller number of variables did help with visualising the complex pathway to high or low lifespan (figure 2).

In addition, because there could be interaction between the variables that is not accounted for, we created another scoring system that considers this. We include this system in the online supplemental material. Considering interaction between variables would increase the precision of the score, but decrease its simplicity. In health and actuarial sciences, scoring systems are often based on diagnostic or pharmaceutical information for the purpose of assessing an individual's disease prognosis or calculating insurance payments. The proposed scoring system is different in that its purpose is to identify groups at high risk of mortality. Nevertheless, including information on health-related behaviours or diagnoses would help make scoring more precise. We did not have such information in the data.

BMJ Open: first published as 10.1136/bmjopen-2023-079534 on 5 August 2024. Downloaded from http://bmjopen.bmj.com/ on September 9, 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

The current article does not aim to predict individual lifespan, rather it aims to provide a tool to better identify individuals at higher risk of mortality. Other studies have attempted to predict individual lifespan, but highlight that, despite considering various predictors, their models could not account for most of the lifespan inequalities.[36 37] Predicting individual lifespan remains a challenging task.

As the PLE between age 30 and 90 years is calculated from period data, our results assume that the age-specific mortality rates observed in 2015–2019 will characterise the life of a 30-year-old until age 90 years from those years. While this is still an ideal measure for summarising population health as it requires minimal data, it might not be ideal for comparing subpopulations defined according to certain characteristics.[25] We take four characteristics and treat them as a snapshot of a series of profiles and assume that these profiles are what an individual experiences for their whole life. This is not a limitation for race or sex (for the most part). We assume changes in educational attainment to be minimal after age 30 years and that this does not have major implications on the results of this study. Other socioeconomic characteristics not considered, such as income and employment status, are more dynamic and could provide better control for economic determinants. These variables were not available in the selected dataset. Marital status is prone to change, and this study only accounted for marital status at a single point in time. By analysing period data, we also do not account for educational and marital status differences across cohorts. For instance, highly educated people from older generations, especially women, represent a selected group, at least partially explaining the magnitude of the inequalities observed in our results. It is also important to note that because mortality has been left truncated at age 30 years, a variable degree of mortality selection has occurred before that age.

Furthermore, we acknowledge that our results are based on unlinked data, used to separately calculate deaths and exposures. This, especially when using education and marital status as indicators of social status, might lead to explaining greater gaps between social groups compared with what would be observed with linked data.[38 39] Finally, survey data are subject to misreporting of information that could lead to biases in the obtained results. This is, for instance, one of the competing causes for observing the Hispanic mortality paradox in the US.[40]

Explanations

Men and women exhibit different behaviours when it comes to health risk and healthcare. For example, women are known to seek healthcare more than men, and have fewer health-risk behaviours such as smoking and alcohol consumption.[41 42] Men are also at higher risk of experiencing violence. In our results, women generally have a survival advantage over men across all causes of death, except for other cancers and Alzheimer's disease. With regard to other cancers, the female disadvantage might be related to female-specific cancers, such as ovarian and uterine cancer, included in this category. It is also well documented that women have a higher incidence of Alzheimer's disease; however, there is evidence that this could be because women have longer life expectancies than men.[43 44]

Since 2000, the US has seen increasing mortality in midlife among non-Hispanic Whites, while mortality rates fell for non-Hispanic Blacks and Hispanics in midlife, as well as those aged 65 years and older across every racial and ethnic group.[45] This has been attributed primarily to increases in mortality caused by drug and alcohol poisonings, suicide, and chronic liver diseases and cirrhosis. Drug poisoning deaths, predominantly by opioid overdose, contributed to a small loss in life expectancy in the US overall.[46] Of the five subpopulations with the lowest life expectancies, Whites accounted for three: never married men with a high school diploma or less, previously married men with a high school diploma or less, and never married women with a high school diploma or less. There are other health risks that differ across racial and ethnic groups. For example, Hispanics tend to have higher rates of high-risk drinking and liver disease,[47] while Whites have higher smoking prevalence.[48] Additionally, Blacks have higher prevalence of obesity,[49] and are more likely to die by homicide. Conversely, despite generally being of lower SES, Hispanics in the US tend to have better health outcomes and lower mortality, a phenomenon known as the "Hispanic mortality paradox".[40] This is attributed to return migration of Mexican migrants, also known as the "salmon bias".

Social factors are known for accumulating their effects throughout the life course of individuals and result in higher risks of developing life-threatening or disabling diseases. This is partially because individuals with different socioeconomic backgrounds have different access to and use of healthcare, different residential and professional arrangements, and exhibit behaviours that might affect health risks. This is certainly captured by education, as individuals with higher degrees have better skills to detect health problems and assess them. Education is also highly correlated with work conditions in adulthood, exposure to risks in the work place, income and health insurance status, leading to overall advantages for those with high levels of education.[50–52]

Historically, mortality in the US has been higher for unmarried people than for married people, and higher still for previously married people in all age groups,[53] including at the oldest ages.[54] It is also commonly reported that marital status is more strongly associated with mortality among men than women, with the latter benefitting less from the protective effect of marriage.[55] For older cohorts, this has often been explained as the effect of women taking care of their husbands at older ages, with a negative effect on women's health.[56] For younger cohorts, although women's roles have changed within households, men still benefit the most from marriage, with their spouses succeeding in positively influencing their health behaviours.[57]

BMJ Open: first published as 10.1136/bmjopen-2023-079534 on 5 August 2024. Downloaded from http://bmjopen.bmj.com/ on September 9, 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Implications

Inequalities in health and lifespan are a burden on society, both socially and economically.[50] These inequalities are associated with a host of outcomes, including unequal access to, and duration of, pensions, among others. The United Nations Department of Economic and Social Affairs highlighted the importance of monitoring and reducing inequalities in lifespan. "Inequality in life expectancy can have a considerable bearing on the length of retirement, running a risk that disadvantaged groups receive less pension as they tend to die earlier. Relatedly, reforms aimed to extend the economic activity and the productivity of the older population by increasing retirement age can be seen as a regressive measure if they do not account for inequality in life expectancy."[58] In addition, the difference in life expectancy between subgroups of the population is costly to a country's economy.[59] Reducing inequalities is vital for societies, and identifying the individuals at high risk of mortality is necessary to succeed.

The scoring system presented in this article can help to identify defined groups with high-mortality risk. For example, public policies cannot target all low-educated individuals as one homogeneous group; we show that low-educated married women have relatively high average lifespans. We also show that the reasons for low lifespan among low-educated individuals vary greatly. For example, smoking-related mortality (eg, lung cancer) tends to be higher among low-educated Whites than low-educated Hispanics. Anti-smoking campaigns targeting individuals of low socioeconomic status could have some degree of impact, but some individuals who are not in risk groups could be targeted, which could lead to a misallocation of resources and inefficiencies in implementation programmes. For example, Hispanic previously married women with a high school degree or less do not have an elevated mortality risk from lung cancer.

The inability to identify a few characteristics responsible for the inequalities in lifespan might complicate the role of public policies. Population-based prevention focuses on determinants which shift the health/mortality distribution for a whole (sub)population.[60] While all the analysed social determinants of health shift lifespan and mortality distributions, they can only explain a small fraction of the inequalities. So, should the focus be more on individual prevention? Rose (2001) argued "Case-centred epidemiology identifies individual susceptibility, but it may fail to identify the underlying causes of incidence".[60] Both types of interventions – individual and population-wide – should not be in competition but complement each other.

The introduced scoring system is a first step in identifying individuals at high risk of mortality. Further studies should be performed including more variables and data from different countries. There is a complex interaction across social determinants of health, but tools exist to simplify and understand it, allowing for better identification of individuals at high risk of mortality.

X Marie-Pier Bergeron-Boucher @bergeron_mp, Julia Callaway @juliacalla and Cosmo Strozza @CosmoStrozza

Acknowledgements We thank James W Vaupel for his guidance and input, and Aleksandrs Aleksandrovs for his help accessing and managing the data.

Contributors M-PB-B and JO designed the research. M-PB-B performed the research. M-PB-B and JO contributed new reagents/analytic tools. M-PB-B, JC and CS analysed the data. M-PB-B, JC and CS wrote the paper. M-PB-B is responsible for the overall content as guarantor. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding The research leading to this publication is part of a project that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No. 884328 – Unequal Lifespans).

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. Data are available in a public, open access repository. The data are publicly available at https://data.census.gov/mdat/#/search?ds=ACSPUMS5Y2019 and https://wonder.cdc.gov/mcd.html.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iDs
Marie-Pier Bergeron-Boucher http://orcid.org/0000-0001-7383-3175
Julia Callaway http://orcid.org/0000-0001-9671-008X
Cosmo Strozza http://orcid.org/0000-0001-8825-2971

REFERENCES

1 Braveman P, Gottlieb L. The social determinants of health: it's time to consider the causes of the causes. *Public Health Rep* 2014;129:19–31.
2 World Health Organization. Social determinants of health. 2022. Available: https://www.who.int/health-topics/social-determinants-of-health
3 Galea S, Tracy M, Hoggatt KJ, *et al*. Estimated deaths attributable to social factors in the United States. *Am J Public Health* 2011;101:1456–65.
4 Lleras-Muney A. The relationship between education and adult mortality in the United States. *Rev Econ Studies* 2005;72:189–221.
5 Marmot M. The influence of income on health: views of an epidemiologist. *Health Affairs* 2002;21:31–46.
6 Brønnum-Hansen H, Baadsgaard M. Widening social inequality in life expectancy in Denmark. A register-based study on social composition and mortality trends for the Danish population. *BMC Public Health* 2012;12:994.
7 Chetty R, Stepner M, Abraham S, *et al*. The association between income and life expectancy in the United States, 2001–2014. *JAMA* 2016;315:1750–66.
8 Steingrímsdóttir OA, Næss Ø, Moe JO, *et al*. Trends in life expectancy by education in Norway 1961–2009. *Eur J Epidemiol* 2012;27:163–71.

BMJ Open: first published as 10.1136/bmjopen-2023-079534 on 5 August 2024. Downloaded from http://bmjopen.bmj.com/ on September 9, 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

9 Barford A, Dorling D, Davey Smith G, *et al*. Life expectancy: women now on top everywhere. *BMJ* 2006;332:808.
10 Beltrán-Sánchez H, Finch CE, Crimmins EM. Twentieth century surge of excess adult male mortality. *Proc Natl Acad Sci U S A* 2015;112:8993–8.
11 Murphy M, Grundy E, Kalogirou S. The increase in marital status differences in mortality up to the oldest age in seven European countries, 1990–99. *Popul Stud* 2007;61:287–98.
12 Dwyer-Lindgren L, Kendrick P, Kelly YO, *et al*. Life expectancy by county, race, and ethnicity in the USA, 2000–19: a systematic analysis of health disparities. *Lancet* 2022;400:25–38.
13 Williams DR, Collins C. US socioeconomic and racial differences in health: patterns and explanations. *Annu Rev Sociol* 1995;21:349–86.
14 Bailey ZD, Krieger N, Agénor M, *et al*. Structural racism and health inequities in the USA: evidence and interventions. *Lancet* 2017;389:1453–63.
15 van Raalte AA, Sasson I, Martikainen P. The case for monitoring life-span inequality. *Science* 2018;362:1002–4.
16 Edwards RD, Tuljapurkar S. Inequality in life spans and a new perspective on mortality convergence across industrialized countries. *Popul Dev Rev* 2005;31:645–74.
17 Shi J, Aburto JM, Martikainen P, *et al*. A distributional approach to measuring lifespan stratification. *Popul Stud* 2023;77:15–33.
18 Vaupel JW, Bergeron-Boucher M-P, Kashnitsky I. Outsurvival as a measure of the inequality of lifespans between two populations. *Demogr Res* 2021;44:853–64.
19 Luy M, Gast K. Do women live longer or do men die earlier? Reflections on the causes of sex differences in life expectancy. *Gerontology* 2014;60:143–53.
20 Fuchs A. Target group. In: Kirch W, ed. *Encyclopedia of Public Health*. Dordrecht, Netherlands: Springer, 2008: 1381–2. Available: https://doi.org/10.1007/978-1-4020-5614-7_3443
21 Bowleg L. The problem with the phrase women and minorities: intersectionality—an important theoretical framework for public health. *Am J Public Health* 2012;102:1267–73.
22 National Vital Statistics System. Multiple Cause of Death File 2015–2019. Hyattsville, MD: National Center for Health Statistics, 2022. Available: https://wonder.cdc.gov/mcd.html
23 United States Census Bureau. American Community Survey (ACS) 5-Year Estimates Public Use Microdata Sample (2019). 2022. Available: https://data.census.gov/mdat/#/search?ds=ACSPUMS5Y2019
24 Xu J, Murphy SL, Kochanek KD, *et al*. *Deaths: Final Data for 2019. National Vital Statistics Reports; Vol. 70 No. 08*; Hyattsville, MD National Center for Health Statistics; 2021. Available: https://www.cdc.gov/nchs/data/nvsr/nvsr70/nvsr70-08-508.pdf
25 Luy M, Di Giulio P, Di Lego V, *et al*. Life expectancy: frequently used, but hardly understood. *Gerontology* 2020;66:95–104.
26 Chiang CL. The Life Table and its Applications. Malabar, FL: Robert E. Krieger Publishing, 1984.
27 Arriaga EE. Measuring and explaining the change in life expectancies. *Demography* 1984;21:83–96.
28 Preston SH, Heuveline P, Guillot M. Demography: Measuring and Modeling Population Processes. Malden, MA: Blackwell Publishers, 2001.
29 Watterberg KL, Aucott S, Benitz WE, *et al*. The Apgar score. *Pediatrics* 2015;136:819–22.
30 Apgar V. A proposal for a new method of evaluation of the newborn infant. *Anesth Analg* 1953;32:260.
31 D'Agostino RB Sr, Vasan RS, Pencina MJ, *et al*. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation* 2008;117:743–53.
32 Wilson PWF, D'Agostino RB, Levy D, *et al*. Prediction of coronary heart disease using risk factor categories. *Circulation* 1998;97:1837–47.
33 Haji Aghajani M, Sistanizad M, Pourhoseingholi A, *et al*. Development of a scoring system for the prediction of in-hospital mortality among COVID-19 patients. *Clin Epidemiol Glob Health* 2021;12:100871.
34 Mishra P, Singh RK, Nath A, *et al*. A novel epidemiological scoring system for the prediction of mortality in COVID-19 patients. *Trans R Soc Trop Med Hyg* 2022;116:409–16.
35 Mehta N, Preston S. Are major behavioral and sociodemographic risk factors for mortality additive or multiplicative in their effects. *Soc Sci Med* 2016;154:93–9.
36 Badolato L, Decter-Frain AG, Irons N, *et al*. Predicting Individual-Level Longevity with Statistical and Machine Learning Methods. Rostock, Germany: Max Planck Institute for Demographic Research, 2023. Available: https://www.demogr.mpg.de/en/publications_databases_6118/publications_1904/mpidr_working_papers/predicting_individual_level_longevity_with_statistical_and_machine_learning_methods_7613
37 Breen CF, Seltzer N. Sociodemographic characteristics alone cannot predict individual-level longevity. *SocArXiv* [Preprint] 2023.
38 Shkolnikov VM, Jasilionis D, Andreev EM, *et al*. Linked versus unlinked estimates of mortality and length of life by education and marital status: evidence from the first record linkage study in Lithuania. *Soc Sci Med* 2007;64:1392–406.
39 Rey G, Rican S, Luce D, *et al*. Measuring social inequalities in cause-specific mortality in France: comparison between linked and unlinked approaches. *Rev Epidemiol Sante Publique* 2013;61:221–31.
40 Palloni A, Arias E. Paradox lost: explaining the Hispanic adult mortality advantage. *Demography* 2004;41:385–415.
41 Preston SH, Wang H. Sex mortality differences in the United States: the role of cohort smoking patterns. *Demography* 2006;43:631–46.
42 White A. Gender differences in the epidemiology of alcohol use and related harms in the United States. *Alcohol Res* 2020;40:01.
43 Fratiglioni L, Viitanen M, von Strauss E, *et al*. Very old women at highest risk of dementia and Alzheimer's disease: incidence data from the Kungsholmen project. *Neurology* 1997;48:132–8.
44 Payami H, Zareparsi S, Montee KR, *et al*. Gender difference in apolipoprotein E-associated risk for familial Alzheimer disease: a possible clue to the higher incidence of Alzheimer disease in women. *Am J Hum Genet* 1996;58:803–11.
45 Case A, Deaton A. Rising morbidity and mortality in midlife among white non-Hispanic Americans in the 21st century. *Proc Natl Acad Sci U S A* 2015;112:15078–83.
46 Dowell D, Arias E, Kochanek K, *et al*. Contribution of opioid-involved poisoning to the change in life expectancy in the United States, 2000–2015. *JAMA* 2017;318:1065–7.
47 Chartier K, Caetano R. Ethnicity and health disparities in alcohol research. *Alcohol Res Health* 2010;33:152–60.
48 Nguyen-Grozavu FT, Pierce JP, Sakuma K-LK, *et al*. Widening disparities in cigarette smoking by race/ethnicity across education level in the United States. *Prev Med* 2020;139:106220.
49 Petersen R, Pan L, Blanck HM. Racial and ethnic disparities in adult obesity in the United States: CDC's tracking to inform state and local action. *Prev Chronic Dis* 2019;16:E46.
50 Cambois E, Brønnum-Hansen H, Hayward M, *et al*. Monitoring social differentials in health expectancies. In: Jagger C, Crimmins EM, Saito Y, eds. *International Handbook of Health Expectancies*. Cham, Switzerland: Springer International Publishing, 2020: 45–66. Available: https://doi.org/10.1007/978-3-030-37668-0_4
51 Montez JK, Hummer RA, Hayward MD. Educational attainment and adult mortality in the United States: a systematic analysis of functional form. *Demography* 2012;49:315–36.
52 Montez JK, Hayward MD. Cumulative childhood adversity, educational attainment, and active life expectancy among U.S. *Demography* 2014;51:413–35.
53 Verbrugge LM. Marital status and health. *J Marriage Fam* 1979;41:267.
54 Manzoli L, Villari P, M Pirone G, *et al*. Marital status and mortality in the elderly: a systematic review and meta-analysis. *Soc Sci Med* 2007;64:77–94.
55 Zueras P, Rutigliano R, Trias-Llimós S. Marital status, living arrangements, and mortality in middle and older age in Europe. *Int J Public Health* 2020;65:627–36.
56 Scafato E, Galluzzo L, Gandin C, *et al*. Marital and cohabitation status as predictors of mortality: a 10-year follow-up of an Italian elderly cohort. *Soc Sci Med* 2008;67:1456–64.
57 Rook KS, August KJ, Stephens MAP, *et al*. When does spousal social control provoke negative reactions in the context of chronic illness? The pivotal role of patients' expectations. *J Soc Pers Relat* 2011;28:772–89.
58 Issifou I, Pewitt J. On the Importance of Monitoring Inequality in Life Expectancy. United Nations Department of Economic and Social Affairs, 2022. Available: https://www.un.org/development/desa/dpad/publication/un-desa-policy-brief-no-145-on-the-importance-of-monitoring-inequality-in-life-expectancy/
59 Asaria M, Doran T, Cookson R. The costs of inequality: whole-population modelling study of lifetime inpatient hospital costs in the English National Health Service by level of neighbourhood deprivation. *J Epidemiol Community Health* 2016;70:990–6.
60 Rose G. Sick individuals and sick populations. *Int J Epidemiol* 2001;30:427–32.

BMJ Open: first published as 10.1136/bmjopen-2023-079534 on 5 August 2024. Downloaded from http://bmjopen.bmj.com/ on September 9, 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.