Using Couples' Data to Estimate the Distribution of Longevity of Older Couples and Surviving Spouses

Janice Compton, University of Manitoba

Robert A. Pollak, Washington University in St. Louis, NBER, and IZA

Seth Sanders, Cornell University

Please Do Not Cite or Quote.

1. Introduction

Spouses may want to know their “joint life expectancy” -- how long they can expect to live as a couple, until "death do us part." And wives and husbands may want to know their “survivor life expectancies” -- how long they can expect to live as widows or widowers if they are surviving spouses.

Implicit in these questions are specifications of the ages of the husband and the wife (e.g., she is 60; he is 62) and the year in which they asked these questions (e.g., 1990). Framed in this way, the questions are about cohort life expectancy for older couples and surviving spouses.

It is tempting to conjecture that we can infer answers to these questions from individuals' life expectancies. Consider a focal couple consisting of a 60 year old non-Hispanic white (NHW) woman married to a 62 year old non-Hispanic white man. Using data on married NHW couples from the National Health Interview Survey (NHIS) from 1986, it is an undergraduate exercise to calculate that

her life expectancy is about 22.0 years and that his life expectancy is about 18.2 years.[1] Intuition suggests that the couple's joint life expectancy is about 18.2 years (the minimum of the spouses' life expectancies), that the wife will be the surviving spouse, and that she can expect to live only a few years as a widow (the difference between the spouses' life expectancies is 3.8 years.)

These intuitions are wrong.

Using individuals' life expectancies to calculate joint and survivor life expectancies yields misleading results, overstating joint life expectancy and dramatically understating survivor life expectancies. Using data on married NHW couples from the NHIS for 1986-1990, we calculate that our focal couple's joint life expectancy is 16.0 years, that if the wife is the surviving spouse, her life expectancy as a widow is 9.8 years, and that if the husband is the surviving spouse, his life expectancy as a widower is 5.2 years. Because some of the 60 year old women and 62 year old men who responded to the NHIS in 1986-1990 are still alive, to calculate these life expectancies requires strong parametric assumptions about age-specific mortalities (e.g., the Gompertz). But with censored data, we can calculate other longevity statistics without strong parametric assumptions such as the median years both spouses are alive, the median years the wife survives after the death of her husband, and the median years the husband survives after the death of his wife. We find that the median years both spouses are alive is 17 years, that if the wife is the surviving spouse, her median years as a widow is 9 years, and that if the husband is the surviving spouse, his median years as a widower is 3 years. This broadly confirms the finding that widows and widowers are likely to live much longer than intuition suggests after the death of their spouse.

---

[1] As we discuss below, relatively small sample sizes in the NHIS require us to use data from the years 1986-1990 and to consider women age 59-61 and men age 61-63. We focus on non-Hispanic white couples because their larger samples in the NHIS allow us to calculate statistics with greater precision using fewer parametric assumptions.

We are not the first to document failures of intuition; see Kahneman (2011). Using the 2010 CDC life tables, Compton and Pollak (2019, 2021) investigate joint and survivor life expectancies by assuming that the life table for all 60 year old NHW women represents the mortality schedule for married 62 year old NHW women as well; and similarly the life table for all 62 year old NHW men represents the mortality schedule for married 62 year old NHW men. Using age-specific mortalities for both men and women from these CDC life tables, they create a joint life table for couples by assuming that at each age (60+*t*) for women and (62+*t*) for men, that the survivor probability of couples is the product of the survivor probability for women multiplied by the survivor probability for men. From this joint life table of couples they can calculate the implied joint and survivor life expectancies. They find that the joint life expectancy is somewhat less than intuition based on individual life expectancies would suggest. They also find that survivor life expectancies are dramatically greater than intuition would suggest and the probability that the husband will be the surviving spouse is substantially greater.

Before turning to the usefulness of a life table approach for estimating joint mortality and the related statistics of couple life expectancy and survivor life expectancy, we review the well-known drawbacks of individual life expectancy from life tables. Most organizations recognize the importance of providing the "right" life table, the principal reason life tables are available separately by sex and race as mortality varies substantially by gender and race. A 60 year old NHW woman in 1990 has at her disposal period life tables by sex and race for many years up to 1990.[2] Most individuals would find the most recent life table for their sex and race and read the calculated life expectancy at their

[2] The CDC has published life tables by race since 1901. Before 1966, these life tables were calculated in Census years when the size of the population by age was measured. Since 1966, annual life tables are published using intercensal population estimates. Life tables by Hispanic origin are available since 2006.

age from the table; in this case she has an expected 23 more years of life.[3] Every student of Demography knows that these tables are period life tables and this calculation is a good predictor of her remaining years to live only if age specific mortality at ages 60 and above have been stable over time. There is of course good evidence that mortality rates have fallen at all ages in the U.S. for a century and most recently a great deal at ages over 60. A more sophisticated individual, say a Demographer, would use multiple period life tables and model the drift in age specific mortality over time, forecast age specific mortality over the next 50 or so years and calculate life expectancy from these forecasts. [4] Of course, the only way to know exactly what she should have expected is in retrospect. Once every NHW woman born in 1930 has died, we can calculate the average number of years lived since 1990 without any assumptions used in various forecast models or in the embedded assumptions in life expectancy calculations. Whatever that number, a specific NHW woman would be unlikely to live to exactly that age as the average has a distribution around it. But we are reassured that at least the average represents the expected value of the number of years she will live (if she had a time machine to travel forward to the year the last NHW women born in 1930 died!)

These concerns are magnified in estimating joint mortality, couple life expectancy and survivor life expectancy, and some new concerns arise as well. To our knowledge, there are no couple life tables; these somehow need to be constructed. The method of Compton and Pollak (2021) attempts to construct these couple life tables from the individual life tables of men and women. This requires (1) that the individual life tables used reflect the mortality schedules of the married men and women modeled; and (2) that the age of death of one spouse is independent of the age of death of the

---

[3] Table 6.3 Vital Statistics of the United States, 1990, Life Tables.

[4] Many methods are proposed to do this but the method of Lee and Carter, 1992, is widely used including by the Social Security Administration to model the size of the Social Security eligible population.

other. Neither of these is likely the case.

Mortality schedules are not available by marital status and so tables on all men and women must be used. But it is likely age specific mortality rates for married women (men) are different than for unmarried women (men). Because, unlike race or sex, marriage is not immutable, we need to be specific about what we mean by "married." The easiest thing would be to define "marriage" so that it was immutable. That is, we could fix marriage at a specific age, say for women at age 60. Then we could answer what is the life expectancy or the median years lived for women who were married at age 60. Below we show that the difference in median life expectancy between NHW women married at age 60 and those unmarried at age 60 is 2.28 years. To put this in perspective, the difference in median life expectancy between all 60 year-old NHW women and all 60 year-old Black women is 1.75 years and between all 60 year-old NHW women and all NHW men is 2.77 years. By these statistics the justification for having (and using) separate mortality schedules by marital status is stronger than having separate mortality schedules by race and almost as strong as having separate mortality schedules by sex.

What we mean by "marriage" for joint mortality is even more vexing. For couples who remain married "until death do us part," one of the partners will (likely) die single – as a widow or widower. In addition, because of "grey divorce" and potentially "grey remarriage," a person married at a point in time may die single or die married but to another partner. What the population is for a "life table for married couples" can be defined differently and would likely be defined by the available data. For example, if marital status is only available at a point in time, the best one could hope to estimate is the longevity of each spouse, regardless of whether they remained married to each other and regardless of remarriages. If we had data on marriage over time we could construct statistics on other

populations, such as the longevity of couples that stay married or the number of years a woman (man) lives past the death of the partner she (he) had at the time marital status was assessed.

We also believe it is unlikely the age of death of one spouse is independent of the age of death of the other in most available data. As noted, it might be possible to construct a mortality schedule and related statistics on longevity for women and men married at a specific ages (say 60 for the woman and 62 for the man). But just as is true for all representative life tables, a specific married woman or man would be unlikely to live to exactly to the age suggested by the average; the statistic has a distribution around it. Her characteristics, like her education or health behaviors, could affect her longevity but so could her husband's characteristics, like his earnings, and vice versa. But again we are reassured that at least across all women (men) married at age 60 (62) the average represents the expected value of the number of years she (he) will live.

The central problem is that if shared characteristics affect individual mortality, they will effect joint mortality and hence couple life expectancy and survivor life expectancy. Consider two situations, one with a strong correlation in mortality between a married woman and man and the other with no correlation. A simple example of strong correlation is that he dies uniformly and randomly over the next 40 years and she dies infinitesimally after him. In this situation, the husband, wife and couple life expectancy are all 20 years and her survivor life expectancy is zero. Now consider a simple example of no correlation, they both die uniformly and randomly over the next 40 years. In this situation, the husband and wife life expectancy remain 20 years and it is quite clear that survivor life expectancy will go up; the wife will now live more than infinitesimally longer than her husband and now the husband will live as a widower as long she would as a widow. In fact survivor life expectancy for each is 13.33 years. Couple life expectancy must fall because time needs to add up. Thinking about

the wife, she expects to live 20 years. If he dies first she expects to live 13.33 years; this will happen 50% of the time so on average she will live 6.66 years without him (50% chance of zero years – he outlives her – and a 50% chance of 13.33 years). This means she that on average she will live 13.33 years with him which is couple life expectancy. Comparing the two examples, survivor life expectancy goes from 0 years to 13.33 years and couple life expectancy falls from 20 years to 13.33 years. In both cases the mortality distribution for men and women are the same - uniform and identical. But the lack of correlation in age at death in the second implies shorter couple life expectancy and longer survivor life expectancy.

A final issue is that even if appropriate marriage specific mortality life tables could be constructed and if by some miracle mortality between spouses was independent, there would still be a problem of forecasting any changes in mortality over time. At best we might construct these tables for a few years; there is not rich historical presence of couple life tables and hence forecasting methods that rely on many period life tables cannot be applied.

Microdata on couples, specifically microdata with the age at death for both the husband and the wife, might help a great deal. In addition to the age at death of both spouses, the ideal micro data set would include information birth cohort, education, age, race, and ethnicity of both spouses as well as information on factors that might cause the mortality of one spouse to be related to the other.

The ideal data set would be longitudinal, following the "marital futures" of the husband and wife, for example, whether the surviving spouse remarried following the end of their marriage by death or whether one or both spouses remarried following the end of their marriage by divorce. Information about the “marital pasts” of both spouses opens the possibility of estimating the correlates of the “protective effects of marriage,” which may depend on the length of the current marriage and

of previous marriages. The analysis of the protective effects of marriage is complicated by the need to distinguish between the effects of selection (healthier people are more likely to marry) and the protective effects of marriage. ((REW or drop this paragraph; or keep for Astrid and Linda Waite.))

Finally, the ideal data set should be large enough to allow non-parametric estimation of the joint distribution of longevities and to do so not only for non-Hispanic white couples, but also for Black couples and Hispanic couples.

Although no U.S. data set satisfies all of these requirements, the National Health Interview Survey (NHIS), an annual cross-section survey that includes both married couples and unmarried individuals, comes closer than any other.[5] Since 1986 the NHIS has been linked to the National Death Index (NDI) and the Social Security Death Index (SSDI) and is considered the "gold standard" (Cowper et al., 2002) for estimating age-specific mortality. The NHIS records the marital status at the time of the survey, for our purposes, 1986-1990, and matches death records to all household members, providing death dates for those who died before the end of 2015. The NHIS also records education for both spouses, allowing researchers to investigate both the effects of an individual's own education on his or her longevity and also the effect of the spouse's education on longevity.

Although the NHIS allows us to estimate joint and survivor life expectancies (e.g., for the cohort of couples who reported in the 1986-1990 surveys that they were married), the NHIS data set is less than ideal. While in some years the NHIS collected some data on "marital past," it did not collect these data in the years for which it provides information on age at death. Because the NHIS is constructed by matching death records to a cross-section survey, "marital futures" data were never

[5] The Health and Retirement Study (HRS) is an alternative data source. Although the HRS has more variables than the NHIS-NDI linked data, the HRS sample sizes are much smaller.

collected. As we stated, the best we can do is estimate the longevity of each spouse, regardless of whether they remained married to each other and regardless of remarriages.

In principle, the NHIS data allow researchers several alternative methods of estimating joint life expectancy (i.e., how long both spouses remained alive) and survivor life expectancy (how long each lived as a widow or widower. If we observed every individual in the NHIS data until all of them had died, then we could estimate any statistic of interest without resorting to parametric methods that depend on assumptions about functional forms (e.g., the Gompertz). We could simply follow individuals and couples over time, counting the number of individuals who died and the number of couples in which one or both spouses died one year after being surveyed, two years after, etc.; couple life expectancy would then be the weighted average of these survival times where the weights are the fraction of couples who "died" after each number of years. Similarly, survivor life expectancy for wives (husbands) could be calculated by making year zero the date of death of a husband (wife) and then counting the number of wives (husbands) dying one year after their husband (wife), two years after, etc.; then taking a weighted average of these survival times, where the weights are the fraction of husbands (wives) who died each number of years after her (his) husband (wife).

Unfortunately, the NHIS only enables researchers to observe deaths that occurred before the end of 2015, 29 years after the initial survey in 1986. This is not equivalent to following both spouses until they die: approximately half of women aged 60 and a quarter of men aged 62 at the time of the survey were alive at the end of 2015 when the link with mortality data ends.[6][7] Finally, the NHIS is not large; we have a working sample of 476 couples. As a result of the limited window of access to

[6] A two year age gap between spouses is typical in these cohorts we investigate.

[7] NCHS periodically links the National Death Index to NCHS surveys. The current link covers deaths through December 31, 2015. A link through December 31, 2019 is expected to be released in Spring 2022.

mortality data and limited sample size, statistics that rely on the entire survival function, such as couple life expectancy, can only be constructed using parametric models such as the Gompertz, Other longevity statistics such as medians can be constructed non-parametrically.

Because of differences in longevity and in sample size by race and ethnicity, fewer statistics can be constructed for Black and Hispanic couples than for non-Hispanic white couples from the NHIS using non-parametric methods.[8]

This paper makes four contributions. The first is to explain why the intuition that joint and survivor life expectancy can be calculated from individual life expectancies is wrong. The key is the randomness of mortality and the overlap in the spouses' mortality distributions. The extreme case is complete overlap and it occurs with same-sex couples in which the spouses are the same age.

The paper's second contribution is to compare the mortality of married women age 60 and married men age 62, with the mortality of various categories of unmarried women and unmarried men of these ages (i.e., the never married, the divorced, and widows and widowers). Using the median age at death as a summary statistic, we find that married individuals live longer. For example, the median age of death for married women (men) is 85.5 years (82.7 years), while the median age of death for divorced women (men) is nearly 3 years (2 years) younger. If life tables reported age at death by marital status, one might think it would be possible to compare the longevity of the married individuals with that of unmarried individuals. For reasons we explain below ((where in the paper??)) using marital status at death yields estimates of the longevity of the first spouse to die. In this paper

[8] For example, since women have greater longevity than men, it is more likely we can estimate median years a widower survives than a widow survives. In addition, because of sample size issues, non-parametric estimates for Black couples or Hispanic couples will be noisier than for non-Hispanic white couples. Therefore the noise reducing properties of parametric models might be necessary to estimate longevity for Black couples and Hispanic couples.

we use marital status when the couple responded to the NHIS during the years 1986-1990 and we usually limit our attention to women (men) who reported their age when they responded to the survey as in the range 59-61 (61-63).

Our third contribution hinges on using the NHIS data on actual couples to construct augmented life tables from which we construct synthetic couples in various ways. We calculate measures of longevity for synthetic couples constructed in various ways and compare them with synthetic couples constructed in other ways and with actual couples. The synthetic couples constructed in this way mimic the characteristics of our population of married individuals more closely than is usually possible because they are based on our sample of married individuals satisfying our time and age inclusion criteria (marital status in 1986-1990; women age 59-61; men age 61-63), and because we control for a wider range of longevity-relevant characteristics (e.g., education.) First, they are constructed from couple data and thus reflect the actual mortality experience of the married individuals we investigate. Second, they are cohort rather the period based. Unlike typical period life tables, we can construct life tables for (a) men and women who, in a particular year or range of years (e.g., 1986-1990) who were married to each other and who during those years satisfied our inclusion criteria (i.e., women age 59-61; men age 61-63) (b) men and women with specific longevity-relevant characteristics (e.g., education). We are able to construct life tables for actual and synthetic couples with a specific age gap between spouses (we focus on couples in which husbands are two years older than wives and both are Non-Hispanic white). And we construct augmented life tables that are conditioned on the education levels of both spouses.

The paper's final contribution is to use the couples' micro data to compare various parametric and nonparametric measures of longevity. We do this because our data are censored – a substantial

number of those who responded to the NHIS in 1986-1990 were alive the end of 2015, so we investigate various measures of longevity that we can construct from censored data. The simplest nonparametric measure is the median. The simplest parametric measure is based on the Gompertz distribution, which allow us to estimate the censored death dates. We also make these comparisons based on older married couples – women age 70 married to men age 72 in 1986-1990 because censoring is much less severe for these older couples.

The ideal measure uses various non-parametric approaches to construct longevity measures. This is possible only with data on actual couples. This allows us to calculate some statistics on couple longevity and survivor longevity without complete mortality data. Because we can construct these statistics with life tables as well, this gives us a way of testing the accuracy of the life table approach.

The paper proceeds as follows. Section 2 constructs and analyzes examples designed to show why the intuition that we can calculate joint and survivor life expectancies from individual life expectancies fails. The analysis then shifts to focus on when the first spouse will die and when the second spouse will die. Once we recognize the overlapping of the spouses' mortality distributions and the randomness of mortality, the need to move from naive intuition to serious analysis becomes clear. Section 3 presents formal definitions and section 4 describes the NHIS data. Section 5 presents NHIS estimates of joint and survivor life expectancy and uses longevity matrices to present a richer and more granular picture of the mortality experience of older Americans. Section 6 is a conclusion.

2. Motivating Examples

2.1 Bob and Carol

Consider Carol, age 60 who in 1986 reported to being married to Bob age 62 who are both non-Hispanic white. Each year, (a) Carol could die and Bob survive, (b) Bob could die and Carol survive, or (c) both Bob and Carol could die. Bob and Carol would like to know how long they can expect to live together as a couple (couple life expectancy) and how many years each can expect to live as a widow or widower if they are the surviving spouse (i.e., the second spouse to die).

Couple life expectancy and survivor life expectancy can be constructed as an extension of standard lifetable methods in a straightforward but tedious way if we had life tables for women and men who in 1986 reported being age 60 and being married. To help Bob and Carol estimate their joint and survivor life expectancies, we could calculate expected transition probabilities to widows, widowers, and "couple death" as each person ages. We would do this by assuming that Carol's mortality rate X years after initial observation would follow the rates of other non-Hispanic women as presented in a lifetable FQ(X); and Bob's mortality rate X years after initial observation would follow the rates of other non-Hispanic men as presented in a lifetable MQ(X).[9] Assuming independence, couple mortality follows from the three possible states of the world (a-c above) and the couple probability of death is calculated as

$$CQ(X) = (1-FQ(X))*MQ(X) + (1-MQ(X))*FQ(X) + FQ(X)*MQ(X). \quad (1)$$

As is standard, the probability of death maps into a mortality rate by taking minus the log of the age specific survival rate, here $CM(X) = -\ln(1-CQ(X))$. When the mortality rate is small $CM(X) \approx -\ln(1-CQ(X))$. Now parallel to standard lifetable construction, we can construct a cohort of 100,000 synthetic couples (rather than 100,000 individuals as is standard) starting the wife at age 60 and the husband at age 62, and using CQ(X) calculate the number of surviving synthetic couples at each age, the number of couple deaths at each age, and from there couple life expectancy using standard lifetable methods. Specifically, couple life expectancy can be

[9] For example, if in 1986 Carol reported she was age 60 and married to Bob who she reported was age 62, when X=10 FQ(X) is her mortality rate at age 70 and MQ(X) is his mortality rate at age 72.

calculated as

$$\sum_{x=0}^{60}\prod_{z=0}^{x}\left(1-CM(z)\right)\approx\sum_{x=0}^{60}\prod_{z=0}^{x}\left(1-CQ(z)\right).^{10} \quad (2)$$

Similarly, we can calculate survivor life expectancy.[11]

If we use the 1986-1990 NHIS to construct a life tables for 60 year old women and for 62 year old men we calculate Carol's life expectancy at age 60 is 24.6 years, Bob's life expectancy at age 62 is 19.9 years but they can only expect to live together as a couple for 16.1 years. There is a 63% chance that Carol will outlive Bob and she can expect to live 13.7 years without him if she does. There is a 37% chance that Bob will outlive Carol and he can expect to live 10.4 years without her if he does. Intuition fails us massively. Intuition suggests that the couple's joint life expectancy is 19.9 years (the minimum of the spouses' life expectancies), rather than 16.1 years and that if Carol is the surviving spouse, she can expect to live 4.7 years without Bob (the difference between the spouses' life expectancies) rather than 13.7 years.

2.2 Alice and Carol

Consider Carol, age 60 married to Alice age 60 who are both non-Hispanic white. Since both women are the same age, both have the same life expectancy, 24.6 years. The fallacy of our intuition now becomes clear; in the case of Bob and Carol, our intuition told us that couple life expectancy was limited to Bob's life expectancy and the difference in life expectancy between the two would give us the number of years that Alice expected to live without Bob. For Carol and Alice, this difference is zero. But if Alice dies first, do we expect that Carol would die immediately and if Carol dies first that Alice would die immediately? Clearly the answer is no, we would expect the surviving spouse to live a number of years as a widow. This shows where our intuition went wrong. In fact, our calculations suggest that Carol and Alice will live 15.2 years as a couple well short of either's life expectancy; and after one dies, the other can expect to live 11.3 years.

2.3 Bob and Ted

Our intuition suggests that when individual life expectancy is lower, couple and

[10] We pick the upper limit to be 60 human longevity is limited to about 122 years, 60 years past when the man is age 62.

[11] Constructed lifetables for the examples below are found in Appendix A.

survivor life expectancy are both lower as well. In fact, the former is lower but the later can be higher. Consider Bob, age 62 married to Ted age 62 who are both non-Hispanic white. Since both are men, both will have the same life expectancy, 19.9 years. Our calculations suggest that Bob and Ted will live 13.2 years as a couple; after one dies, the other can expect to live 12.1 years. The lower couple life expectancy for Ted and Bob vs. Carol and Alice stems from higher age specific mortality rates because they are men and because they are two years older than Carol and Alice. But notice that survivor life expectancy is longer if Bob predeceases Ted than if Carol predeceases Alice.

3 Which life table should we use, or can we use any at all?

Rewriting Equation 1, helps focus on which life table we should use and whether the life table approach can be used at all. Equation 1 can be rewritten as,

$$\mathrm{CQ(X)} = \mathrm{MQ(X)} + \mathrm{FQ(X)} - \mathrm{FQ(X)}*\mathrm{MQ(X)}. \tag{3}$$

Substituting Equation 3 into Equation 2, we get couple life expectancy can be calculated as

$$\sum_{x=0}^{60}\prod_{z=0}^{x}\big(1-CQ(z)\big) = \sum_{x=0}^{60}\prod_{z}^{x}\Big(1-\big(\mathrm{MQ(z)}+\ \mathrm{FQ(z)}-\ \mathrm{FQ(z)}*\mathrm{MQ(z)}\big)\Big)$$

$$= \textstyle\sum_{x=0}^{60}\prod_{z=0}^{x}\Big(\big(1-\mathrm{MQ(z)}\big)+\big(1-\mathrm{FQ(z)}\big)-\ \big(1+\ \mathrm{FQ(z)}*\mathrm{MQ(z)}\big)\Big) \tag{4}$$

To build intuition, suppose that the husband never died, that is, $\mathrm{MQ(z)} = 0\ \forall\ z$. Then couple life expectancy would reduce to

$$\textstyle\sum_{x=0}^{60}\prod_{z=0}^{x}\big(1-\mathrm{FQ(z)}\big) \tag{5}$$

which is the definition of female life expectancy; and similarly if the wife never died couple life expectancy would equal husband life expectancy. This make clear that for a couple like Carol and Bob, whatever lifetable is used needs to represents the individual mortality schedule for both Carol and for Bob (and their individual life expectancies); in addition, Equation 4 suggests there needs to be sufficient conditioning information so that their joint probability of death can be modeled as the product of their individual mortality rates (conditional independence).

Often lifetables are available by sex and race but rarely are they available by marital status. These life tables reflect mortality rates calculated across marital status and partner characteristics. If mortality rates vary by marital status, then the appropriate lifetable would be ones constructed on

married men and married women. Even if lifetables for married men and married women were available, they may not be sufficient when considering a specific case (Carol age 60 married to Bob age 62) if mortality rates vary with spouses' age or other characteristics. Further, even if lifetables for married men and women by spouse characteristics (such as age) were available, this may still not be sufficient to calculate couple life expectancy if unobserved factors cause the conditional independence assumption to fail. For example, lifetables are not available by health behaviors and these are likely correlated for husbands and wives. Finally, the limitations of using period lifetables to trace out cohort life expectancy for individuals is well known and equally apply to couple cohorts.

When we use a typical lifetable constructed by sex and race, what fraction are representative of Carol's marriage to Bob. Table 1 answers this using the 1990 Census. In Panel A, Columns 1 and 2 calculate statistics for Non-Hispanic white women and men. We see that only 70% of NHW women aged 60 are married. Many are widowed (15.1%) and the remaining are single because of divorce (11.2%) or because they never married (3.5%). Of the women who are married, what fraction, are married to men exactly like Bob - that is, men age 62 who are also non-Hispanic white. Panel B shows that only 13.0% of are married to men exactly age 62 (Carol is two years younger than Bob; even allowing a broader age range to represent that the husband is "around 2 years older," say the wife is one to three years younger, this is only 36% of marriages among 60 year old NHW women. Race and Ethnicity play less of a role; 97.4% of married 60 year old non-Hispanic white women are married to non-Hispanic white men. Overall, we estimate that of non-Hispanic women represented in a lifetable available only by sex and race, only 9% are married to non-Hispanic men exactly age 62; and even if we allow a broader definition of the husband is "around 2 years older" only 24.9% are married to non-Hispanic men ages 61-63. Panel B shows a similar story for 62 year old non-Hispanic white men. The vast majority are not married to women exactly two years younger (only 12% are) and only 33.9% are married to women "around 2 years younger." Overall, only 9.9% are married to non-Hispanic women exactly age 60; and even if we allow a broader definition of the wife is "around 2 years younger" only 27.6% are married to non-Hispanic women exactly ages 59-61.

If Carol and Bob were either Black or Hispanic, the mortality statistics in the life table would be even less influenced by marriages like theirs. Only 40% (61.9%) of Black women (men) age 60 (age 62) are married. And only 9.3% (8.3%) of Black women (men) are married to men (women) two

years older (younger). Again using a three year age window, only 26.1% (25.1%) of Black women (men) are married to men (women) "around two years" older (younger). Statistics for Hispanic men and women are similar to Black men and women.

That lifetables average mortality rates over many subpopulations of people of a specific race and gender matters only if mortality rates vary across these subpopulations. We use the 1986-1990 NHIS matched to death certificates to investigate this. Because the NHIS is substantially smaller than the 1990 Census, we analyze women age 59-61 and men ages 61-63 to approximate women age 60 and men age 62 (hereafter referred to as "Age 60" and "Age 62.") Table 2 shows that mortality rates vary substantially by marital status for both men and women. The median age of death for married non-Hispanic white women age 60 at the time of the survey is 85.4 years. The median age of death for women who were divorced (82.4 year), widowed (83.3 years) or never married (83.6 years) at the time of the survey is several years lower. Non-Hispanic men display a similar pattern of married men having greater longevity than men who are not married as to other Black and Hispanic men and women.

Figure 1a shows that even within marriage, own mortality rates vary substantially by the age of one's spouse. The highest median age at death for married non-Hispanic white women age 60 at the time of the survey, is 85.7 years, for Black women it is 86.2 years and for Hispanic women it is 88.5 years. For NHW women, the highest median age of death occurs when she is one year older than her husband; for Black women the highest median age of death occurs when she is two years older than her husband. But the median age of death changes very little around the peak suggesting that the highest median ages of death occur when NHW and Black women are about the same age as their husbands. The highest median age of death for Hispanic women is 88.2 years. This occurs when she is 10 years younger than her husband and her median age at death drops steeply as she ages relative to her husband.

Figure 1b shows parallel statistics for married men age 62. The highest median age at death for married NHW men is 83.2 years; for Black men it is 80.9 years and for Hispanic men it is 86.4 years. Again how the median age at death changes as a function of spouses age parallels the graph for women. Generally, the maximum median age at death for NHW and Black men occurs when they are 3 or 4 years older than their wives. But for Hispanic men, the maximum occurs when they are 10

years older than their wives and the median age at death falls as they are younger relative to their wives. We note that while the majority of men and women are within 6 years of their spouse, there is still a substantial number of couples with more than a 6 year difference in age between spouses.

Finally, Figure 2 addresses whether mortality of husbands and wives are likely independent. To do this we analyzed the age of death of every husband of women age 59-61 when surveyed in the NHIS. Ninety-five present of husbands range in age from 54 to 79 with a median age around age 62 and with 80% of husbands between the age of 58 and 68. We estimated a flexible Gompertz hazard model where both the base hazard rate and the effect of aging were allowed to be a flexible function (cubic) in age by race. We then calculated the median age of death for men of each age at the time of the survey and classified each man as living beyond what the model predicts is the median age of death for his race, or, dying before that median age. Having classified each husband, we then ask whether the survival function of wives is effected by whether her husband lived beyond the median age of death. Figure 2 shows the strong relationship between when a wife dies and when her husband dies. Wives that had husbands that lived longer lived substantially longer themselves. For example, the median age of death for wives who husbands lived longer than their median age at death was 88; the median age of death for wives who husbands died before their median age at death was 85.

What this suggests is that typical lifetables represent some types of couples better than others; it also suggests that the conditional independence assumption used in Compton and Pollack (2021) likely does not hold. It is an empirical matter how flawed are calculations that use typical lifetables and maintain the conditional independence assumption. We turn to this below.

4. Data

The National Health Interview Survey (NHIS) is the principal source of information on the health of the civilian noninstitutionalized population of the United States. The main objective of the NHIS is to monitor the health of the United States population through the collection and analysis of a broad range of health related data. The 1986-2011 NHIS surveys are matched to the NDI as of December 31, 2015. Importantly, the NDI is matched to <u>all</u> household members. This makes the NHIS a unique source for our work because it records the age of death for both a husband and his wife. We use the 1986-90 NHIS, which allows the up to 29 years to observe deaths of

survey respondents.. Because of data limitations, we focus on legal marriage and ignore cohabitation. A second major advantage of the NHIS is that it contains detailed demographic characteristics as well as health status and health behaviors of respondents at the time of the survey.

While the NHIS has central advantages, it also has drawbacks. As we will show, even with a 29 year follow-up, about half of 60 year old women remain alive and about a quarter of 62 year old men remain alive as of December 31 2015. Second, because the NHIS is a sample, we generally will aggregate across several years of age; for example above we aggregated women 59-61 to represent women "about age 60." Given the limited sample size and the incomplete recording of ages of death due to censoring, we implement parametric models for much of our work. Specifically we use parametric models to calculate all statistics that rely on estimating the complete distribution of age of death (such as couple life expectancy and survivor life expectancy). However, we show that other statistics need not rely on the complete distribution of age of death and these we compute non-parametrically from our data.

The fourth contribution exploits the NHIS data on education and on the age gaps between husbands and wives. Using life tables constructed from NHIS data on couples, we construct synthetic couples and investigate how closely their longevity matches the longevity of actual couples with the same or similar education and age gap. We find that the joint and survivor life expectancies are very similar, but that the pattern in the tails of the longevity matrix differ. Both the upper tail and the lower tail of the longevity matrix for actual couples contain more couples than the longevity matrix of the synthetic couples. That is, compared with the synthetic couples, we find more actual couples in which both spouses die earlier than predicted and more actual couples in which both spouses die later than predicted by our simulation. This finding is consistent with marriage (usually) having a (causal) protective effect and with marriage having a selection effect that operates through assortative marriage. The protective effect on longevity operates through shared health-related behaviors such as diet and exercise and access to medical care (e.g., health insurance) which are not reported in the NHIS. The effects of marriage are not necessarily protective, although the fact that married individuals live longer than unmarried individuals is consistent with the hypothesis that, on average, marriage is protective. The assortative marriage hypothesis is that healthy people marry healthy

people and unhealthy people marry unhealthy people. This assortative marriage hypothesis is a selection effect, but who marries whom is distinct from selection into marriage – the claim that health people are more likely to marry than unhealthy people. Selection into marriage is consistent with random matching in the marriage market and this is not what we observe. We think it likely that both selection into marriage and assortative marriage play roles, but we have no strategy for estimating their relative importance.

Table 1:
Characteristics of Women Age 60 and Men Age 62
Census of Population and Households, 1990

Panel A: All Men and Women

	(1) NHW Women Age 60 (Percent)	(2) NHW Men Age 62 (Percent)	(3) Black Women Age 60 (Percent)	(4) Black Men Age 62 (Percent)	(5) Hispanic Women Age 60 (Percent)	(6) Hispanic Men Age 62 (Percent)
Marital Status						
Married	70.2	82.8	40.0	61.9	59.4	76.3
Divorced/Separated	11.2	8.9	24.9	21.9	17.8	15.0
Widowed	15.1	3.7	27.4	8.2	17.3	4.1
Never Married	3.5	4.6	7.7	8.0	5.6	4.5
	.	.	.	.	.	.
Total	50429	46327	5351	3952	1811	1416

Panel B: Married Men and Women (Spouse Present)

Age Difference						
Six or More Years Younger	21.6	2.9	30.3	3.8	25.4	4.3
Five Years Younger	7.8	1.0	7.3	1.6	7.2	1.3
four Years Younger	9.2	1.3	7.8	1.6	6.7	0.8
Three Years Younger	11.6	2.3	7.9	2.2	8.7	1.7
Two Years Younger	13.0	3.3	9.3	3.4	10.1	3.6
One Year Younger	11.4	5.5	8.9	5.3	8.6	5.1
Same Age	10.1	9.5	8.6	7.0	8.7	7.6
One Year Older	5.7	10.7	5.2	8.5	6.1	9.8
Two Years Older	2.8	12.0	3.9	8.3	4.1	10.3
Three Years Older	1.8	11.2	2.7	9.0	2.4	9.5
Four Years Older	1.1	8.9	1.4	8.3	2.5	8.6
Five Years Older	0.8	7.1	1.8	7.1	2.6	5.8
Six or More Years Older	2.9	24.4	4.9	33.9	6.9	31.6
Race/Ethnicity Same						
Yes	99.1	98.6	98.9	96.6	87.4	88.0
Married to "Right" Person						
Yes Exactly	9.0	9.9	3.8	4.9	5.7	7.1
Yes Approximately	24.9	27.6	10.0	14.7	14.8	20.8
Total	35417	38336	2140	2445	1075	1081

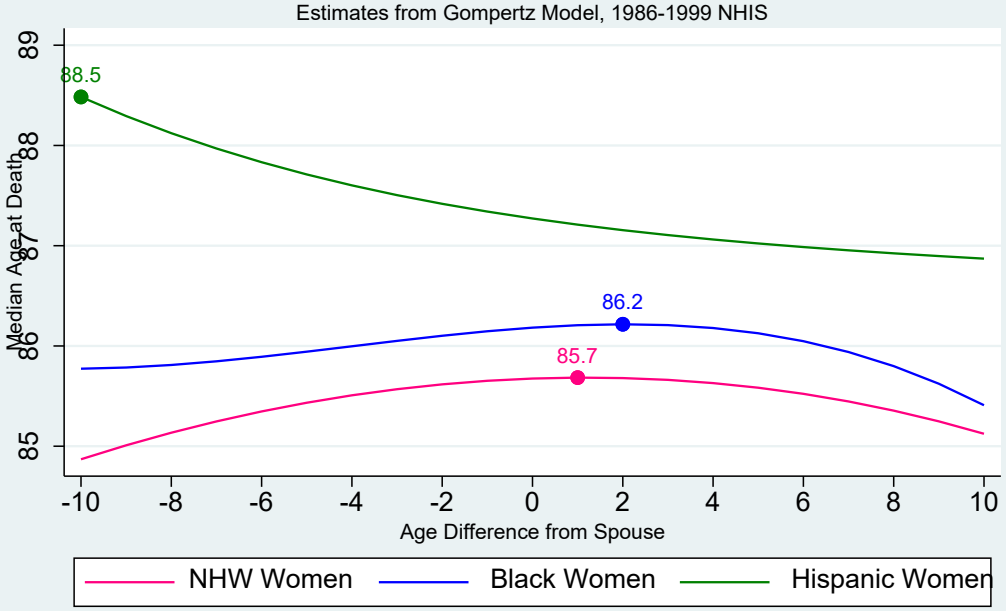
Table 2:
Median Age at Death, Women Age 59-61 and Men Age 61-63
National Health Interview Survey, 1986-1990

Panel A: All Men and Women

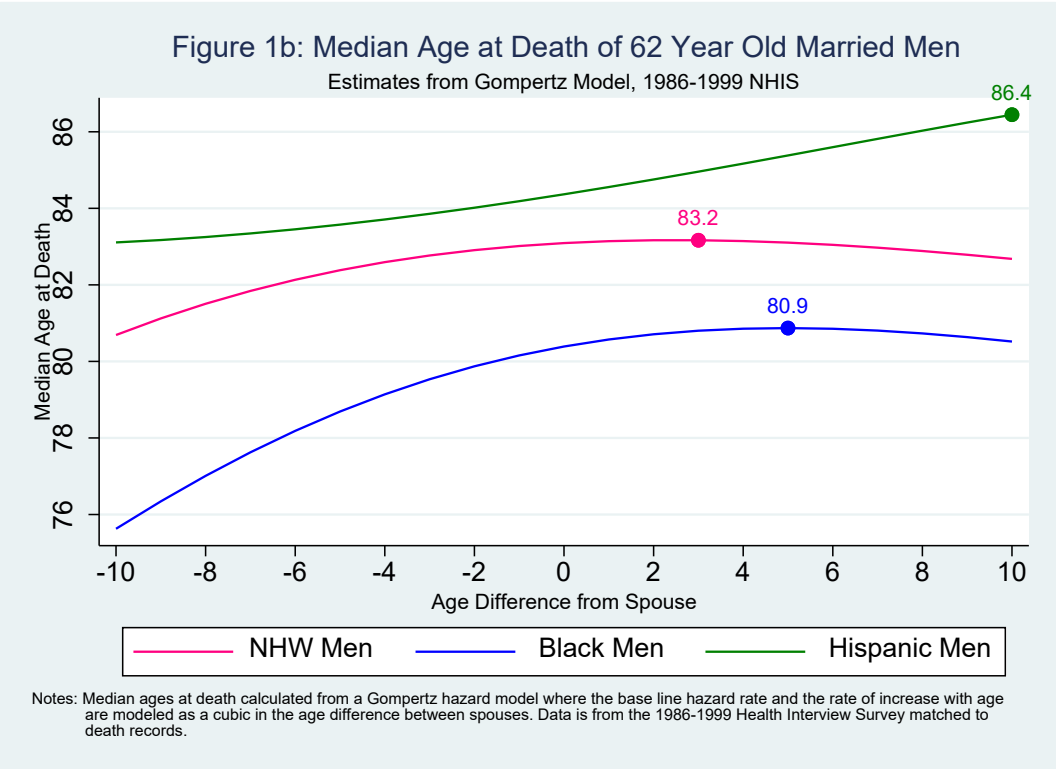
Marital Status	(1) Women Age 59-61 Median Age at Death	(2) Men Age 61-63 Median Age at Death	(3) Black Women Age 59-61 Median Age at Death	(4) Black Men Age 61-63 Median Age at Death	(5) Hispanic Women Age 59-61 Median Age at Death	(6) Hispanic Men Age 61-63 Median Age at Death
Married	85.4	82.8	85.2	80.2	87.8	85.3
Divorced/Separated	82.4	80.4	83.6	80.1	85.4	81.8
Widowed	83.3	79.5	82.3	78.3	83.4	84.4
Never Married	83.6	80.7	81.6	76.1	87.5	82.5
Total	12,471	13,985	2,612	1,865	1,160	1,480

Figure 1a: Median Age at Death of 60 Year Old Married Women

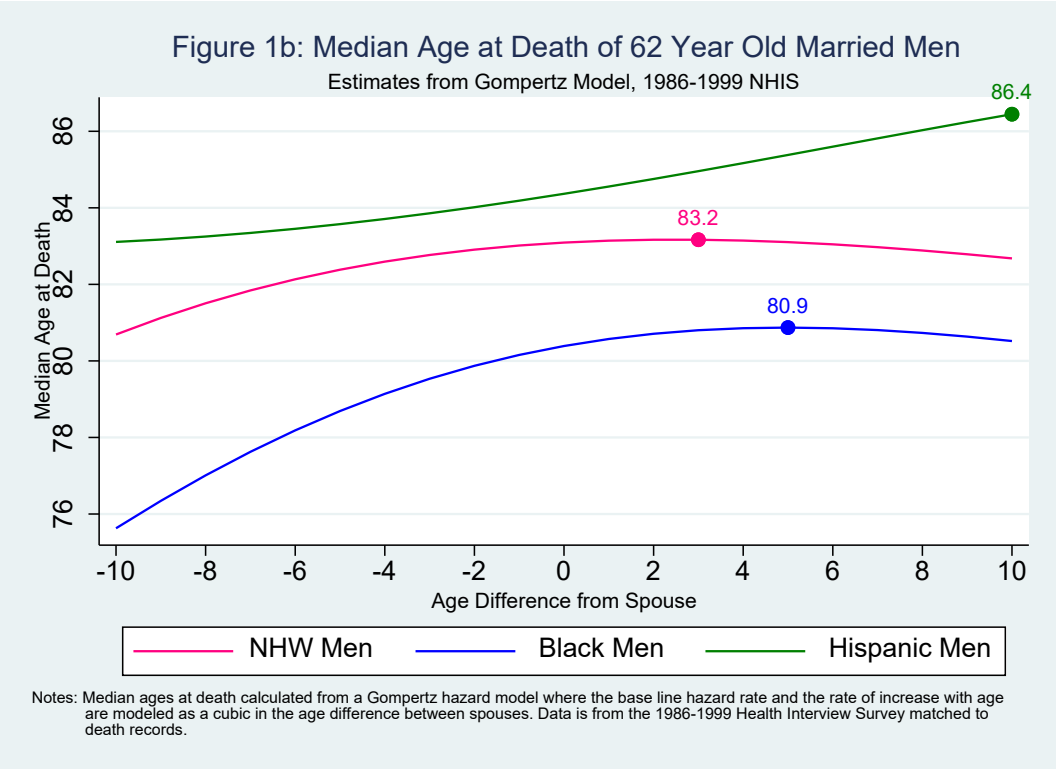
Estimates from Gompertz Model, 1986-1999 NHIS



Notes: Median ages at death calculated from a Gompertz hazard model where the base line hazard rate and the rate of increase with age are modeled as a cubic in the age difference between spouses. Data is from the 1986-1999 Health Interview Survey matched to death records.



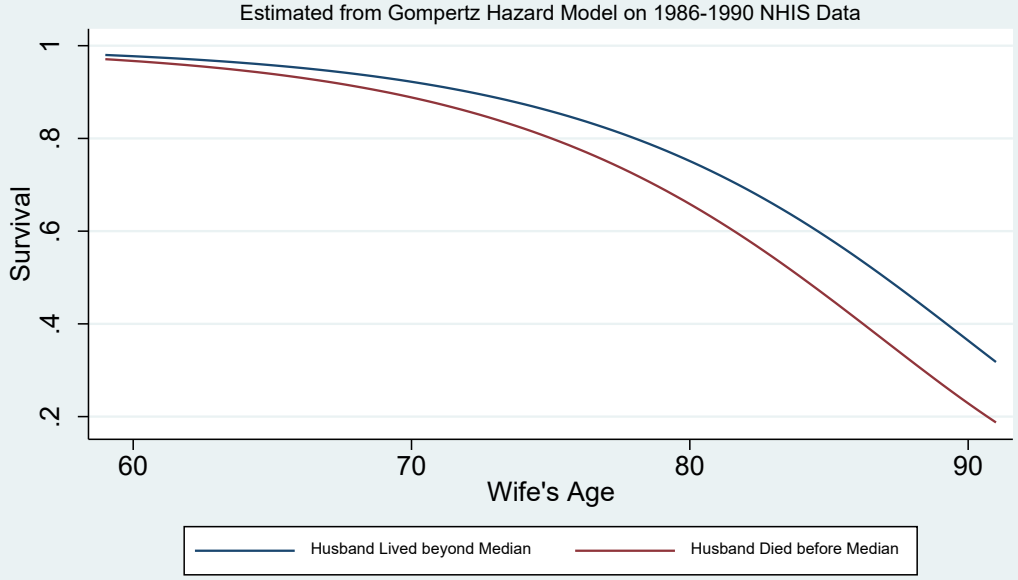
Notes: Median ages at death calculated from a Gompertz hazard model where the base line hazard rate and the rate of increase with age are modeled as a cubic in the age difference between spouses. Data is from the 1986-1999 Health Interview Survey matched to death records.



Notes: Median ages at death calculated from a Gompertz hazard model where the base line hazard rate and the rate of increase with age are modeled as a cubic in the age difference between spouses. Data is from the 1986-1999 Health Interview Survey matched to death records.

Figure 2: Survival of Married Women by Death Age of Husband

Estimated from Gompertz Hazard Model on 1986-1990 NHIS Data



Notes: Husband was classified as living beyond the median age for married men his age at the survey as follows - a Gompertz hazard model where the base line hazard rate and the rate of increase with age is modeled as a cubic in the husband's age at the survey. Then the median age of death for husbands of each age is predicted. Husbands living past the median age for his race are classified as "Lived beyond Median." If husband died before median age he is classified as "Died Before Median." Husbands last observed alive but at an age below median age of death were droped (1,632 of 11,657 observations).

These calculations assume independence in mortality rates between the husband and his wife, an assumption that is necessary if data on actual couples is unavailable. The probability that both survive from ages a_j to $(a+1)_j$ where $j \in (w,h)$ is one minus the sum of the three probabilities above.

Our estimates differ from Compton and Pollak (2021) only in that we construct life tables for married men and women from the NHIS rather than use the CDC lifetables that are not available by marital status. When life expectancy for 60 year old women is calculated using only the married female mortality table we find life expectancy is 24.3; for 62 year old men using only the married male mortality table we find life expectancy is 18.8. But when we calculate couple life expectancy from the extended life table we find it is 15.5 years and survival life expectancy of a wife is 13.1 years (and the probability she has a 34% chance of predeceased her husband!)[12] If she predeceases her husband survival life expectancy of the husband is 9.8 year. How can it be that couple life expectancy is lower than the life expectancy for either men alone or women alone? How can it be that survival life expectancy for the wife is 13.1 years where our intuition tells us it is only 5.5 years?

[12] Appendix A provides these calculations when the NHIS data is used to construct life tables for married women approximately age 60 and married men approximately age 62. Compton and Pollak (2021) provide similar calculations using mortality tables for all non-Hispanic white men age 62 and non-Hispanic white women age 60.