Why the Married Live Longer:

Evidence on the Selection versus Protection Effect of Marriage using Fixed-effect Survival Models

Iliana Kohler* Hans-Peter Kohler† Axel Skytthe‡

September 23, 2011

Abstract

A robust finding in the demographic research is that married people have a lower risk of death as compared to divorced or single individuals. Two main hypotheses have been proposed to explain this relationship between marriage and mortality: marital selection versus marital protection. In this paper we investigate this interdependence between marriage and mortality proposing new statistical methods and data. We develop fixed-effect survival model that to our knowledge has not been applied to the analysis of mortality and twin data. This fixed-effect survival model allows the estimation of the parameter of interest, which measures the impact of individual characteristics on the level of mortality, without imposing the assumption of independence between the distribution of unobserved heterogeneity and individual characteristics. Our analyses are based on the Danish twin register. We compare the estimates between MZ and DZ twins. While both share same socioeconomic background during childhood, MZ twins share also the same genetic determinants of mortality. Our results suggest that the effect of marriage on mortality may be primarily traced back to the effect of selection rather than to marital protection.

1 Introduction

Since the pioneering work of William Farr (1858), it is well documented that health varies significantly across marital status categories. This issue has recently been the topic of the cover story of a *New York Times Magazine* cover article (Parker-Pope 2010). Numerous studies, for instance, have documented that married individuals face a considerably lower risk of death than divorced and widowed people, or those remaining single (e.g., Burgoa et al. 1998; Durkheim 1951; Hu and Goldman 1990; Lillard and Panis 1996; Morgan 1980; Murray 2000; Preston and Taubman 1994; Rendall et al. 2011; Trovato and Lauris 1989; Welon et al. 1999). In addition, substantial benefits of being married have been shown for a large spectrum of chronic diseases, functioning problems and disabilities (Kravdal 2001; Pienta et al. 2000). Despite the fact that many aspects of marriage—including the onset and duration, the relations among married partners, the processes of household decision-making and work allocation, the tasks performed within and outside the

*Associate Director, Population Aging Center, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104; *Email:* iliana@pop.upenn.edu.

†Frederik J. Warren Professor of Demography, Department of Sociology and Population Studies Center, University of Pennsylvania, Philadelphia, PA; *Email:* hpkohler@pop.upenn.edu.

‡Senior data analyst, University of Southern Denmark, Odense, Denmark; *Email:* askytthe@health.sdu.dk

household, etc.—differ in important ways across socioeconomic strata, cultures, countries or time periods, the health and mortality gradient by marital status has been remarkably persistent and prevails almost universally in the contexts where it has been investigated (e.g., Hu and Goldman 1990). Marriage is also beneficial for both husband and wife, despite their considerably different social and economic roles within households and marriages (e.g., Jacobs and Gerson 2004).

Despite this overwhelming evidence that marriage is associated with improved health and lower mortality, the underlying mechanism and in particular the causal relations between marriage and health/mortality outcomes are not fully resolved. In particular, there is no clear evidence that the "marriage advantage" in mortality or health can be attributed to the functions performed by marriage—such as increased social and emotional support, mutual insurance, resource pooling, social control, or other household services—rather than to the characteristics of the individuals who get married. Marriage may just happen to be associated with better health and lower mortality due to positive selection into marriage, that is, due to the fact that healthier individuals are more likely to get and remain married. However, even the reverse relation is conceivable. For instance, less healthy individuals may devote more effort to finding a partner and remaining married because these individuals enjoy substantially larger marriage gains than healthy individuals; as a consequence, analyses may also underestimate the benefits of marriage because individuals are negatively selected (e.g., see Lillard and Panis 1996).

The complexity of the relationship between marriage and mortality has resulted in two main explanations for the beneficial effects of marriage on health and mortality: *marital protection* versus *marital selection*. The first explanation emphasizes the socially integrative function of marriage and relates the association between marriage and lower risk of death to protective mechanisms operating through the economic settings, social environment and health behaviors characterizing marital unions. In particular, the long-term commitment associated with marriage allows the husband and wife to enjoy benefits that are not fully reaped in less stable unions, such as for instance the gains from specialization, economies of scale in household production, mutual insurance, and long-term support. (Trovato and Lauris 1989; Waite 1995; Zick and Smith 1991). In addition, through social control mechanisms in the marital unions, married men and women engage in less risky behaviors and are more likely to adopt healthy lifestyles (Tucker and Anders 2001; Umberson 1987). Several authors have also argued that marriage leads to positive health outcomes including mental health by providing greater societal integration through larger social networks and relationships (Durkheim 1951; Gove 1973; Horwitz et al. 1996; House et al. 1988; Rutledge et al. 2003; Simon 2002; Waite and Lehrer 2003).

In summary, the explanation that marital unions are protective for adverse health and mortality outcomes is derived from multiple factors that characterize the social environment of the married individuals and tend to promote health and well-being. However, problems of selection bias affect many studies since married individuals are not a random subset of the population, but rather a selected one, and marital partners are not assigned to each other randomly but assortatively with respect to health, socioeconomic characteristics, preferences, etc. The alternative explanation therefore argues that it is not the fact of being in a marital union, but rather the observed and unobserved characteristics of individuals who enter and remain in marital unions that

explain the health and mortality advantage associated marriage. The marital selection hypothesis therefore emphasizes the role of selection with respect to health, socioeconomic and genetic characteristics occurring in the marriage market as the underlying explanation for the health and mortality advantage of married men and women (see for example Goldman 1993; Hu and Goldman 1990; Kisker and Goldman 1987; Lillard and Panis 1996; Murray 2000). While some studies claim that these selection arguments cannot explain the health and mortality advantages associated with marriage, as for instance Waite and Gallagher (2000) who state that "[t]he selection of happy and healthy people into marriage cannot explain the big advantage in mental and emotional health husbands and wives enjoy" (p. 68), the empirical basis for these conclusions is often weak. For instance, studies have included observed characteristics of the marriage partners in their analyses to control for the selection into marriage, which leaves open the possibility that married individuals are importantly selected with respect to unobserved characteristics—including both unobserved social backgrounds, such as conditions of the parental household, or genetic dispositions. Moreover, survival (or event-history) models with unobserved frailty, which have also been proposed as a solution to address the selection problem, rely on important assumptions, including (*i*) assumptions about the correct specifications of the relation modeling the selection into marriage, and/or (*ii*) independence assumptions between the explanatory variable included in the analyses, of which the most important in the context of this paper is the marital status, and the unobserved individual characteristics that are represented through the "frailty" component of the model.

Despite the apparently ample evidence on the association between marriage and mortality, therefore, the causal links underlying these associations is still subject to a heated debate. In particular, most existing studies of the marriage-health-mortality relations are not able to distinguish between causal and spurious effects by applying standard methods of analysis that do not appropriately account for unobserved heterogeneity and selection into marriage on basis of individual's endowments like family backgrounds, social contexts in youth and genetic dispositions. In this paper, we therefore take an alternative and new approach in order to disentangle the causal relationship between marriage and mortality. In particular, we develop fixed-effect survival models that can overcome many of the identification problems when combined with suitable sibling or twin data. In the method we propose here, the estimation of the parameter of interest, which measures the impact of individual characteristics on the level of mortality, does not require the assumption of independence between the distribution of unobserved heterogeneity and the individual characteristics considered in the analysis as standard survival methods that include unobserved frailty usually do.

This fixed effect approach is particularly powerful when applied to data on monozygotic twins that have the potential of controlling for both unobserved social and genetic endowments. Because monozygotic (identical) twins share the same genetic endowment as well as the same parental background and various social or economic endowments related to the parental household, we can use fixed effect analyses within monozygotic (identical) twins in order to control for a wide range of unobserved factors—including genetic dispositions and early-life contexts associated with the parental household—that affect both health and mortality outcomes as well as marriage propensity. A similar approach has been used extensively in the analysis of the returns to educa-

tion (e.g. Ashenfelter and Krueger 1994; Ashenfelter and Rouse 1998; Behrman and Rosenzweig 1999; Behrman et al. 1996; Behrman and Taubman 1976) or household allocations and marriage market effects (Behrman and Rosenzweig 2002, 2004; Behrman et al. 1994; Conley et al. 2003), and controlling for endowments has often substantially changed the estimates and therefore the conclusions in these studies.

The remainder of this paper is organized as follows. In the next section we discuss the problem of unobserved heterogeneity in mortality analyses and develop the fixed-effect survival model as a new approach to this issue. In Section 2, we present the twin data used in our analysis that are based on the Danish twin register, and in Section 4 we discuss the results obtained from the estimation of the fixed-effect frailty model and compare these results to standard survival and random-effect survival models. We also compare the results from fixed-effect model to estimates obtained from behavioral-genetic analyses on marriage and mortality, and we conslude in the final section with a discussion of the future potential and alternative applications of the approach proposed in this paper.

2 Data

Twins studies long have been used to assess the multiple roles of endowments on demographic and socioeconomic outcomes and on estimates of effects of various variables net of such endowments in both the psychological/behavior-genetic and the socioeconomic literatures (for a recent review, see Kohler et al. 2011). Our analyses are based on the Danish Twin Registry that was established in 1954 as the first nationwide twin registry in the world (see Hauge 1981; Hauge et al. 1968; Kyvik et al. 1996, 1995). This registry covers twins born during the period 1870–1982. Data from this twin registry have extensively been used for analyses of health, mortality and aging (e.g., see Christensen 2001), psychological phenotypes (McGue and Christensen 1997, 2001) and fertility (Kohler and Rodgers 2003; Kohler et al. 1999).

We restrict our analyses to the old twin cohorts born between 1880–1940, who were alive on April 2nd 1968 when the Danish Civil Registration System *(CRS)* was implemented in the country.[1] In addition, for the purpose of our analyses, we select only same-sexed MZ and DZ twins with verified zygosity.[2] We restrict our sample to complete twin pairs for which we know the survival status of both twins. That is, we exclude twin pairs in which one or both twins have emigrated or live outside of Denmark, or who were lost during the period of observation. Our analyses are also conditional on survival of both twins to age 50. The summary statistics of the data used in the present analyses is given in Table 1.

The marital history of the twins is potentially relevant for our analyses. However, one of the limitations in our dataset is that the Danish twin registry provides exact information on changes of the twins' marital status only after April 2nd 1968, when the Civil Registration System was introduced in the country. As we focus on the old twin cohorts born 1880–1940, most of the

[1]The Danish Civil Registration System (CRS) encompasses all persons who have lived in Denmark since 2 April 1968 and have registered with the national registration offices.

[2]The verification of the zygosity is based on a survey including four questions about the similarity of the twins, and this method has been proved to determine the zygosity correctly in approximately 95% of the twin pairs Hauge 1981.

Table 1: Summary statistics of the Danish twin data used for the analysis of mortality differentials by marital status.

	MZ Twins		DZ Twins	
	Males	*Females*	*Males*	*Females*
Total	1342	1043	2692	2016
Alive	525	469	1137	969
Dead	817	574	1555	1047
	In complete twin pairs with at least one death		*In complete twin pairs with at least one death*	
Total	880	576	1770	1102
Alive	112	61	306	165
Dead	768	515	1464	937

changes in their marital history (if any at all), may have occurred prior the implementation of the CRS in the country. Because we are not able to obtain the complete marital history of the twins in our sample, we use in our analyses marital status at the time, when the first death occurs, or for twin pairs without a death, the current marital status.

The preliminary analyses in this PAA submission are based on register data that include both marital histories and vital status until the year 2000. In the meantime, we have obtained additional data from the Danish Twin Registry that includes marital histories and mortality until 2008, providing 8 years of additional follow-up as compared to the analyses presented below. These additional data will strengthen the statistical power of our analyses, using these data, we will also be better able to investigate some of the puzzles that remain in our preliminary analyses. For instance, the analyses presented below show a reduced mortality of divorced individuals. Also, due to an insufficient number of observed deaths, our current analyses cannot identify changes in the effect of marriage on mortality by age, or interactions of the current marital status with indicators of the marital history (eg., total number of years married, etc). We expect that the additional years included in the follow until 2008, which have has only very recently become available for analyses, will allow us to extend and substantially strengthen our existing analyses.

3 The Fixed-Effect Survival Approach as an Alternative to the Standard Survival Analysis

The consideration of unobserved heterogeneity is a central issue in the assessment of the *marital selection* versus *marital protection* hypotheses, and more general, in the assessment of how individual behavioral decisions (e.g., marriage, smoking or drinking) impinge upon health and mortality outcomes. In the context of survival models, unobserved heterogeneity has been primarily incorporated via gamma-distributed relative frailty models (Vaupel et al. 1979; Yashin and Iachine 1995), or simultaneous models for marital status, health and mortality (Lillard 1993; Lillard and Panis 1996), and related random-effect survival (or event-history) models. In the following we will argue that these models incorporate unobserved heterogeneity only on the basis of strong parametric assumptions that are not necessarily innocuous for the interpretation of the estimation

results. Most importantly, these models are not able to overcome one main concern associated with the effects of unobserved heterogeneity: the correlation between unobserved heterogeneity and observed behavioral determinants of mortality, as for instance in our application, the correlation between marital status and unobserved health status. The fixed-effect survival model we develop in the following sections allow us to overcome many of these problems when combined with suitable data, such as for instance data on mortality of twins. In particular, we show the estimation of the fixed-effect survival model when the observed individual characteristics are not time-varying and we expand the model further so that individual time-varying characteristics can be also incorporated in the analyses. Furthermore, in the appendix to this paper we show how under certain assumptions, the fixed-effect survival model can be extended so that the residual individual-specific heterogeneity observed between monozygotic twins can be considered and modelled.

Before turning to the development of this model and its extensions, we first present the related problem of random and fixed-effect estimation in linear regression and we briefly review the prevailing approach to the incorporation of frailty in mortality models via random effect models.

3.1 Background: random versus fixed effects in linear regression

The common problem that unobserved heterogeneity can be potentially correlated with important explanatory variables used to explain demographic or economic outcomes, has received considerable attention in the literature and applications of linear regression models (for related discussions, see for instance Behrman et al. 2011, 1994; Kohler et al. 2011). For illustration of this problem, consider the simple linear regression model

$$y_{it} = x_{it}\beta + u_i + e_{it}, \tag{1}$$

where y_{it} is an outcome variable observed for person i at time t and x_{it} is a set of explanatory variables of person i at time t. The residual of this regression represents unobserved determinants of the outcome y_{it} and can be decomposed in two components: the fixed effect u_i characterizes persistent unobserved heterogeneity among individuals, such as for instance differences in health status or abilities, while the second term e_{it} reflects contemporaneous effects on the outcome at time t.

Two dominant approaches exist for the estimation of this type of models. If both u_i and e_{it} are independent of the explanatory variables x_{it}, then *random effect models* can be applied to Eq. (1) in order to obtain a consistent and efficient estimates of the parameters of interest in β. However, the independence assumption is likely to be violated in many cases. For instance, if the variables on the right-hand-side include marital status and y_{it} is a measure of mortality, then unobserved characteristics contain in u_i, such as health, are likely to be correlated with x_{it}. In this case, random effect models no longer provide an unbiased estimate of β. The alternative method that can be implemented despite the correlation of u_i and x_{it} is the *fixed effect estimation*. This method relies on the much weaker assumption that only the contemporaneous residual influence, e_{it}, is independent of the explanatory variables in x_{it}. The estimation of the model is then possible if we observe

within-individual variations of the outcome and explanatory variables over time, which allows sweeping out the unobserved individual characteristics in u_i.In this case, unbiased estimates of β are obtained even when random effect models fail to do so.

Unfortunately, fixed effect estimation has not been implemented in the context of survival models and many existing approaches in the literature on mortality remain subject to the criticism associated with the random effect model above: the correlation between x_{it} and u_i can lead to distorted inferences of how the explanatory variables in x_{it} affect the outcome variable y_{it}.

3.2 Unobserved heterogeneity in the random-effect mortality models

Consider the case where the hazard of mortality of a person at age a is given by

$$\mu(a|x,z) = h(a) \cdot z \cdot \exp(x\beta), \tag{2}$$

where z represents unobserved individual characteristics called in our application frailty, and x is a vector of—for simplicity time-invariant—factors influencing the risk of death for the person at age a. The probability density function for the age at death, conditional on frailty and the observed characteristics of a person, is then given by

$$f(a|z,x) = \mu(a|x,z) \cdot s(a|x,z), \tag{3}$$

where $s(a|x,z) = \exp[-\int_0^a \mu(\alpha|x,z)d\alpha]$ is the probability of surviving to age a.

The main problem here arises from the fact that the density in Eq. (3) cannot be used in the estimation since frailty z is unobserved. In order to overcome this estimation problem, random-effect survival models assume a specific parametric distribution of the unobserved characteristics z and then integrate out the unobserved term. In the case when we assume that z has a gamma distribution with mean one and variance σ^2, the respective solution for the population density at age a is particularly simple and is given by the following equation (Vaupel et al. 1979):

$$\begin{aligned} f(a|x) &= \int f(a|z,x) g_\sigma(z) dz & (4) \\ &= \mu(a|x,z=1)[1-\sigma^2 \log s(a|x,z=1)]^{-(1+1/\sigma^2)}, & (5) \end{aligned}$$

where $g_\sigma(z)$ denotes the assumed gamma p.d.f. for unobserved frailty with mean one and variance σ^2. The key assumption in the derivation of the above result is that the gamma distribution of unobserved frailty, $g_\sigma(z)$, does not depend on the explanatory variables in x. That is, the mean and variance of the frailty-distribution is independent of the characteristics contained in x (the same assumption is necessary in alternative random-effect models that assume different distributions for the unobserved frailty). In spirit, this assumption is equivalent to the assumption in the linear regression case, where random effect models assume that the distribution of the individual-specific effect u_i is independent of the right-hand-side variables included in the regression.

This assumption, however, may not necessarily hold, and especially, it may not hold when the effect of behavioral variables on mortality is to be evaluated. For instance, in the context where marriage is selective with respect to health and frailty, married individuals will have a different

distribution of unobserved frailty than unmarried respondents. In this case, random effect models may not yield a correct inference of how marital status relates to mortality outcomes.

3.3 Fixed-effect survival models as an alternative approach

Fixed-effect survival models can overcome the above distributional assumption. However, this approach has not been applied in the context of mortality to the authors best knowledge. One central problem is that the differencing of observations over time, as feasible in the linear regression model in Eq. (1) with multiple observations per individual, is not possible in the context of survival models (simply stated—individuals die only once). This issue can, however, be overcome when more than one outcome is observed that is affected by the same unobserved characteristics. This setting is observed for twins, in particular, for monozygotic (MZ) or identical twins that share the same genetic endowments relevant for mortality outcomes as well as the same socioeconomic background during childhood.

In the following we will therefore consider the mortality and socioeconomic characteristics of twin i in MZ twin pair j. In particular, we assume that for monozygotic twins the relation

$$\mu_{ij}(a) = h(a) \cdot z_j \cdot \exp(x_{ij}\beta) \tag{6}$$

holds, where z_j is the twin-pair specific frailty that is shared by twin 1 and twin 2 within the MZ pair j. For simplicity, we consider again only time-invariant explanatory variables x_{ij}; the corresponding extension to time-varying explanatory variables is straightforward. The survival probability of twin i to age a that is implied by the hazard in Eq. (6) then follows as

$$s_{ij}(a) = \exp[-z_j \cdot \exp(x_{ij}\beta) \cdot \int_0^a h(\alpha)d\alpha]. \tag{7}$$

The key problem in estimating the survival model in Eqs. (6) and (7), *without* assuming the independence of z_j and x_{ij}, is to find an equivalent procedure to the differencing in the linear regression model that sweeps out the unobserved characteristics from the likelihood function.

Fortunately, this can be achieved via a conditional likelihood approach. First, denote as A_j^F the age when the first death in twin pair j occurs. The density of this first death is then given by

$$f(A_j^F | x_{1j}, x_{2j}, z_j) = [\mu_{1j}(a) + \mu_{2j}(a)] \cdot s_j^p(a),$$

where $s_j^p(a) = s_{1j}(a) \cdot s_{2j}(a)$ denotes the probability that both twins in pair j survive to age a, $\mu_{ij}(a)$ is the mortality hazard of twin i in pair j at age a. According to Eqs. (6) and (7), both $\mu_{ij}(a)$ and $s_{ij}(a)$ depend on the frailty z_j and the individual characteristics x_{ij}.

Second, denote as D_j^{F1} a dummy variable that equals one if twin 1 is the first to die within pair j and zero if otherwise. Then consider the probability that twin 1 in pair j is the first member of the twin pair to die, conditional on the age A_j^F when the first death in twin pair j occurs. This

conditional probability follows as

$$\begin{aligned}\Pr(D_j^{F1}=1|x_{1j},x_{2j},A_j^F) &= \frac{\mu_{1j}(a)}{\mu_{1j}(a)+\mu_{2j}(a)} \\ &= \frac{\exp(x_{1j}\beta)}{\exp(x_{1j}\beta)+\exp(x_{2j}\beta)} \\ &= \frac{1}{1+\exp(-\Delta x_j\beta)} \end{aligned} \tag{8}$$

where $\Delta x_j = x_{1j} - x_{2j}$ is the difference in the socioeconomic mortality determinants between both twins within pair j.

The key insight obtained from Eq. (8) is that the conditional probability $\Pr(D_j^{F1}=1|x_{1j},x_{2j},A_j^F)$ is independent of the unobserved individual heterogeneity z_j. This unobserved frailty is swept out be considering which twin is the first to die within twin pair j conditional on the age at first death A_j^F within pair j. Equation (8) therefore allows a fixed-effect approach to survival data that does not require the assumption that unobserved frailty is independent of the behavioral variables that determine the mortality level. (For a general theoretical discussion of conditional maximum likelihood estimation, see Gourieroux and Monfort 1995; a common example of conditional MLE estimation is the fixed-effect logit model, see Andersen 1970 and Hsiao 1986).

The estimation of the relevant parameter β in Eq. (8) is particularly simple since it is obtained from a logistic regression of the dummy variable D_j^{F1} on the within-twin pair difference Δx_j in the mortality determinants. Moreover, despite the fact that the parameter β is estimated via a logistic regression, it has the standard proportional-hazard interpretation that $\exp(x_{ij}\beta)$ is the relative change in the mortality hazard associated with the individual characteristics x_{ij}.

3.3.1 Time-varying individual characteristics

In our above model, we have assumed that the observed characteristics x_{ij} that determine the level of mortality are time-invariant and therefore fixed over the respondent's life. However, we know that many individual-level characteristics change over time. Thus, an extension of the above approach to time-varying characteristics is necessary, and as we see below this is extension is particularly straightforward. Let us denote as $x_{ij}(a)$ the observed characteristics for twin i in pair j at age a and denote as $X_{ij}(a) = \{x_{ij}(0),...,x_{ij}(a)\}$ the complete history of these characteristics from age zero to age a. To avoid a cluttering of the notation, we frequently drop the reference to the age a in our notation of the history X_{ij} when the age-reference is clear from the given context.

While the mortality hazard $\mu_{ij}(a) = h(a)\cdot z_j \cdot \exp(x_{ij}(a)\beta)$ at age a depends only on the contemporaneous characteristics at age a, the survival probability to age a depends on the complete history and is given as

$$s_{ij}(a|X_{ij}) = \exp[-z_j \int_0^a h(\alpha)\exp(x_{ij}(\alpha)\beta)d\alpha].$$

Despite these potentially different history of individual characteristics, the fixed-effect estimator of β remains unaffected and continues to be given by Eq. (8), where the only modification is that the

within-twin pair difference in the explanatory variables is calculated at the age of first death with the twin pair; that is, the difference Δx_j in Eq. (8) is replaced with $\Delta x_j(A_j^F) = x_{1j}(A_j^F) - x_{2j}(A_j^F)$.

Our random-effect model relies on the assumption that both twins share the same unobserved characteristics and both twins are identical in terms of frailty. However, we know that even though MZ twins are genetically identical at conception and share the same socioeconomic background in childhood, they may differ regarding some unobserved individual-specific characteristics that influence their risk of death. Thus, in the appendix of the paper, we develop the model further and show how potential differences in the twins' characteristics can be considered and incorporated in the analyses when relaxing some of the underlying assumptions.

4 The Effect of Marriage on Mortality Revisited

In this section we begin our empirical analyses with a focus on the relationship between marriage and mortality. In standard regression analyses, the estimation of the marriage effect on health/mortality and in particular the direction of causality (i.e. marital protection versus selection into marriage) is generally obstructed by problems caused by unobserved endowments that affect simultaneously the left-side variable—in our case, mortality—and the explanatory variables, in the present analyses, current marital status. Two primary approaches exist to overcome the estimation problems caused by unobserved endowments in regression analyses. First, we can assume that unobserved characteristics are either not correlated with the observed outcomes, or alternatively, their distribution is independent from the observed characteristics used as explanatory variables (i.e., the latter estimation is known as random-effect models with unobserved heterogeneity/frailty). The second approach uses fixed effect estimations that difference out common (unobserved) determinants of behaviors or outcomes over time (in the case of multiple observations over time) or across individuals (in the case when individuals share common endowments). The comparison between the two approaches is of a particular importance as it allows testing the independence assumptions that underlie standard survival and random-effect models. If unobserved endowments and frailty are random with respect to marriage so that there are no differences in the distribution of unobserved characteristics between married and single individuals, than the fixed-effect survival model is expected to yield similar results to the ones obtained by the random-effect and standard survival models. That is, the trends in the mortality differentials by marital status estimated from these different models would be expected to conform with each other. Alternatively, if the models yield different results, this indicates that the right-side variables are correlated with the disturbance term, in which the unobserved endowments are embedded. Thus, the independence assumptions underlying standard and random effect models are distorted and we should be cautious, in particular when interpreting causal relationships derived from these letter models.

Because of the above reasons, in this section we pursue the strategy of comparing the estimates obtained from the fixed-effect, standard and random-effect survival models and we discuss to which extent differences in the results impinge on the interpretation of the causal relationship between marriage and mortality. In particular, we estimate a standard piecewise-constant expo-

nential model (i.e., a non-parametric estimation of the baseline hazard of mortality) and a standard Gompertz model (i.e., a parametric estimation in which the mortality hazard increases exponentially with age). In addition, we estimate both models as random-effect models with unobserved heterogeneity/frailty. In these letter models, frailty is assumed to be gamma-distributed with a mean 1 and a variance σ^2. All models are estimated separately by sex and for MZ and DZ twins 50-95 years old. The risk of death is held constant within 5-year age intervals.[3] The piecewise-constant and the Gompertz models include also dummies for 10-years cohorts (not reported in the results). Table 2 shows the results from the various estimations in the form of coefficients and standard errors in parenthesis.

Model 1 in Table 2 reveals for men, the 'classical' pattern of mortality differentials by marital status known from several other studies that have looked at the beneficial effects of marriage for health outcomes. Married MZ twins experience the lowest risk of death among men, while single MZ twins have about 61 per cent $(\exp(.479) = 1.61)$, and divorced MZ twins have 57 per cent higher relative risk of death as compared to married MZ twins.[4] For DZ male twins these differences between matrimonial categories are even stronger. According to the non-parametric estimation of the risk of death, single DZ male twins have almost twice as higher risk of death as married DZ male twins. The magnitude of the mortality differentials between married and divorced DZ twins is comparable to the respective estimates for MZ twins. In contrast to the estimates for MZ male twins, widowed DZ twins have about 46% higher risk of death than married men, while this difference is not statistically significant for MZ twins. When we consider the effect of unobserved heterogeneity in our model (model 1 with frailty), the magnitude and path of the above patterns of mortality differentials by marital status do not change and only the difference in the risk of death between married and single MZ and DZ twins increases.

In contrast to male twins who seem to follow the 'classical example' of mortality differentials by marital status, the estimates for female twins are not straightforward and differ from the male pattern. In particular, we estimate that only single MZ and DZ female twins have statistically significant higher risk of death compared to the reference group of married women. While in the standard nonparametric approach, the risk of death estimated for single women ranges from 31% (DZ twins) to 33% (MZ twins) compared to married women, this differences increases to about 40% if we consider the effect of unobserved heterogeneity in our estimation. In contrast to male twins, the mortality differentials between divorced, widowed and married women are not statistically significant for both MZ and DZ twins, and this pattern does not change if we include frailty in our estimation. However, it is noticeable that the coefficients for divorced and widowed female MZ and DZ twins have a negative sign, which insinuates that women in these two marital categories may even experience a lower risk of death than their married counterparts.

Model 2 in Table 2 shows the parametric estimation of the mortality differentials by marital status in which the baseline hazard of mortality is assumed to follow a Gompertz hazard function. Similarly to Model 1, in addition to the standard model we estimate also a random-effect Gompertz model which considers the effect of unobserved heterogeneity. A comparison between

[3] We do not report the baseline hazard of mortality. However, both estimations—the parametric and the non-parametric—highly agree with the standard mortality patterns (i.e., the risk of death increases with age).

[4] The relative risk of death is calculated as $\exp(\beta)$.

Table 2: Estimation coefficients of fixed-effect, random effect and standard survival models for MZ and DZ Danish twins born between 1880–1940, who have survived above age 50.

	Monozygotic Twins				Dizygotic Twins			
	Males		Females		Males		Females	
Model 1—Piecewise-constant survival model[a]								
		with frailty		with frailty		with frailty		with frailty
Single	0.479	0.653	0.291	0.326	0.646	0.739	0.274	0.317
	(0.132)**	(0.144)**	(0.128)*	(0.148)*	(0.084)**	(0.086)**	(0.098)**	(0.105)**
Divorced	0.451	0.434	-0.138	-0.223	0.431	0.453	-0.015	-0.020
	(0.169)**	(0.202)*	(0.194)	(0.219)	(0.121)**	(0.125)**	(0.119)	(0.130)
Widowed	0.011	0.051	-0.088	-0.098	0.381	0.388	-0.026	-0.034
	(0.115)	(0.140)	(0.119)	(0.126)	(0.082)**	(0.091)**	(0.083)	(0.087)
Variance of	–	0.422	–	0.302	–	0.153	–	0.135
frailty σ		(0.092)		(0.106)		(0.049)		(0.067)
Model 2—Gompertz survival model[b]								
		with frailty		with frailty		with frailty		with frailty
Constant	-8.881	-10.736	-10.022	-11.514	-9.671	-10.522	-9.964	-10.588
	(0.453)**	(0.646)**	(0.540)**	(0.715)**	(0.326)**	(0.407)**	(0.404)**	(0.483)**
$b-$parameter	0.084	0.108	0.091	0.111	0.087	0.098	0.092	0.100
	(0.005)**	(0.007)**	(0.006)**	(0.008)**	(0.003)**	(0.004)**	(0.004)**	(0.006)**
Single	0.476	0.646	0.293	0.323	0.649	0.756	0.268	0.308
	(0.133)**	(0.141)**	(0.128)*	(0.146)*	(0.086)**	(0.086)**	(0.098)**	(0.104)**
Divorced	0.460	0.455	-0.136	-0.209	0.431	0.455	-0.023	-0.028
	(0.169)**	(0.199)*	(0.192)	(0.215)	(0.123)**	(0.126)**	(0.120)	(0.129)
Widowed	0.014	0.075	-0.098	-0.110	0.369	0.387	-0.035	-0.046
	(0.112)	(0.134)	(0.119)	(0.123)	(0.083)**	(0.090)**	(0.083)	(0.086)
Variance of	–	0.353	–	0.274	–	0.161	–	0.123
frailty σ		(0.086)		(0.098)		(0.046)		(0.059)
Model 3—Fixed-effect survival model								
Single	0.316		0.127		0.627		0.362	
	(0.275)		(0.305)		(0.164)**		$(0.207)^{+}$	
Divorced	-0.124		-1.205		0.223		-0.271	
	(0.383)		(0.419)**		(0.256)		(0.250)	
Widowed	-0.348		-0.434		0.382		-0.413	
	(0.318)		$(0.253)^{+}$		$(0.225)^{+}$		(0.186)*	

Notes: Standard errors in parentheses. *p-values:* $^{+}$ $p < 0.10$; * $p < 0.05$; ** $p < 0.01$. *a*, *b*: the standards errors are adjusted for clustering within twins pairs by using the Huber-White estimator of variance. The piecewise-constant and the Gompertz survival models include dummies for cohorts (not reported in the table). The models with frailty assume gamma-distributed frailty with mean 1 and variance σ^2.

the nonparametric estimation in Model 1 and the Gompertz model reveals remarkable consistency between the two estimations that remains even when unobserved heterogeneity is controlled for. Moreover, the magnitude and sign of the coefficients estimated for the different marital groups and both fraternal and identical twins do not change substantially in the second model.

In summary, the piecewise-constant hazard and the Gompertz parametric models yield results that conform to the 'classical' gradient of mortality differentials by marital status: married individuals experience the lowest risk of death compared to singles, or divorced and widowed men and women. The pattern is consistent for both, identical and fraternal twins, and for the estimation with and without unobserved heterogeneity. The latter does not change the patterns, but only increases the magnitude of the mortality differentials by marital status. While the mortality gradient by marital status is clearly evident for men, the results for women are not straightforward and with the exception of single women who have a higher risk of death compared to the reference category (married women), the difference in the mortality hazard of divorced and widowed women and married females is not statistically significant in the standard survival models. The above analyses based on standard survival models provide evidence that being married is beneficial for mortality outcomes and this suggests that marriage has a protective effect for health.

The effect of unobserved heterogeneity in Model and Model 2 is modeled under the presumption that frailty is independent of the observed individual-specific characteristics. This assumption is not innocuous as for instance, single and divorced men and women may differ in mortality relevant unobserved characteristics because such characteristics may be correlated wit the propensity to marry or divorce. The fixed-effect survival model proposed by the authors allows us to circumvent this assumption by differencing out twin-pair specific unobserved factors. Thus, it can be applied as a 'control' model for the consistency of the assumptions in the standard survival approaches discussed above.

Table 2 shows that the estimates from the fixed-effect survival model are consistent with the previous analyses only for DZ twins. Similarly to the standard approaches, the new model estimates that married DZ men have the lowest risk of death, while single men experience 87% and widowers 47% higher risk of death compared to the reference category. The positive effect of being divorced remains for DZ male twins but is not statistically significant in the fixed-effect survival model. For females, we find a very similar pattern as the one observed in the standard survival estimations. Single DZ female twins are characterized by the highest risk of death among women that is almost half as higher as among married women. The coefficients for divorced and widowed females remain negative, however, the difference in the risk of death between widows and married women becomes statistically significant in this model.

In contrast to the estimates for fraternal twins, the fixed-effect model yields for MZ twins strikingly different results compared to the standard survival approaches discussed above. The previous results based on the standard survival methods are partially reversed and the 'classical' pattern of mortality differentials by marital status looses completely its significance for MZ male twins. For men there remains a positive effect of being single on mortality, which is smaller than the effect found in our earlier models, and is not longer statistically significant. More striking is the fact that the coefficient estimates for divorced and widowed MZ male twins become negative

indicating that previously married men may be a subject to a lower risk of death than married. Albeit this effect is not statistically significant, it is nevertheless, important because it is in striking contrast to the earlier estimates from Model 1 and Model 2 revealing a considerably higher mortality for divorced and widowed men. Moreover, similar pattern is estimated also for female MZ twins, for whom the point estimates are negative and become substantially larger in magnitude, and the lower mortality of both divorced and widowed women, emerges now as a statistically significant fact.

In summary, the standard and random-effect survival models on one hand and the fixed-effect model on the other hand yield strikingly different estimates of how mortality patterns differ by marital status in particular identical twins. This contrast in the estimates indicates that the independence assumption is violated and in fact, unobserved characteristics are highly correlated with the propensity to get married and respectively with the observed outcome 'mortality by marital status'.

4.1 Marital selection or marital protection—what can we learn from twin studies?

The comparison between MZ and DZ twins within the fixed-effect survival approach is of a particular importance to understand the relationship between marriage and mortality, and especially the causal mechanism behind the beneficial effect of marriage on health and mortality outcomes. As discussed earlier in this paper, several studies report that marriage has a positive effect on health and is associated with a lower risk of death for men and women, which suggests that marital protection may explain the causal mechanism behind this association. However, problems of selection bias affect most of these studies since married individuals are not a random subset of the population, but rather a selected one on the basis of observed and unobserved characteristics, and marital partners are not assigned to each other randomly but assortatively with respect to health, socioeconomic status, preferences, family backgrounds, etc. In addition, an essential problem in the existing research is that none of the studies controls for potentially important biological and family endowments that may determine both, the propensity to get married as well as the health of the individuals.

An alternative approach to account for the above problems of selectivity and endowment effects is to use a twin-design for the analyses. In particular in this paper, we utilize the fact that MZ and DZ twins share the same parental background and various social or economic endowments related to the parental household. Moreover beyond this, monozygotic twins share also the same genetic endowments as they are 100% genetically identical whereas DZ twins share the genetic code of 'normal' siblings (i.e., 50% genetically identically). We can therefore use fixed effect analyses within twin pairs in order to control for a wide range of unobserved factors that affect both the chances on the marriage market and mortality. In particular, the *within-MZ twin pair analyses* can eliminate the influence of unobserved endowments resulting from genetic dispositions and shared environments resulting from the parental households (the vast majority of MZ twins in our data grow up together) and other common socioeconomic contexts (e.g., cohort influences). A similar approach has been used extensively in the analysis of the returns to education (e.g., Ashenfelter and Krueger 1994; Ashenfelter and Rouse 1998; Behrman and Rosenzweig 1999; Behrman

et al. 1996; Behrman and Taubman 1976) or household allocations and marriage market effects (Behrman and Rosenzweig 2002, 2004; Behrman et al. 1994; Conley et al. 2003), and controlling for endowments has often substantially changed the estimates and therefore the conclusions in these studies. Moreover, the fixed-effect approach and the within-twin pair analyses allow us to disentangle the causal mechanism behind the relationship between marriage and mortality. For instance, if marriage influences health and mortality primarily through social, economic, cultural and parental household factors observed in marital unions, then the fixed-effect approach will yield similar results and similar patterns of mortality differential by marital status for identical and fraternal twins (again, both MZ and DZ twins share similar socioeconomic and parental background). In contrast, if marriage is associated with unobserved selective factors that operated in the marriage market and determine the probability to find a partner, then our model will yield different estimates for MZ and DZ twins. The letter suggest that the causal relationship between marriage and health/mortality can be explained by selective mechanisms.

Table 2 shows that the estimates are indeed strikingly different for identical and fraternal twins. While the 'classical' pattern of mortality differentials by marital status is still observed for DZ twins (in particular DZ male twins), it looses completely its significance for MZ male twins. For females, the fixed-effect approach estimates that previously married MZ twins may even experience lower risk of death than those currently married. The differences in the mortality patterns observed between MZ and DZ twins suggest that the effect of marriage is indeed associated with unobserved selective mechanisms. In addition, the results also indicate that standard survival analyses may overestimate the effect of being married on mortality. The effect for females is rather surprising, but the authors think that it is consistent wit the marital selection explanation and discuss this further in the conclusions.

Our analyses suggest that the beneficial effect of marriage on mortality may be explained by the role of unobserved selective endowments rather than by marital protection. However, we do not know anything about the nature of these unobserved endowments, e.g. whether these are biological/genetic or social endowments. Even though we assume that MZ and DZ twins share same social and parental background, there may essential differences in the social endowments due to the fact that MZ twins may be socialized and nurtured in a different way compared to DZ twins. Thus, we pursue our analyses further and apply a behavioral genetic model to illustrate the relevance of biological and social endowments for variation in ever being married. In particular, the twins design allows a decomposition of the within-population variation in ever being married across individuals into three components: *a)* variance that is consistent with influences of genetic factors; *b)* variance that is consistent with shared environmental factors; and *c)* variance that is consistent with individual-specific influences that are not common to both twins within a pair. It is important to note, despite interpretations that often are given in the literature, that these variance decompositions do *not* reveal causality. If, for example, genetic heritability is high, that does *not* mean that environmental changes might not have important causal effects. Our motives in presenting such estimates are not to permit inferences about causality, but to situate our study in the literature and to explore whether there are likely biases in estimates that attempt to identify the causal impact of marriage on health and mortality without controlling for genetic and other

Table 3: ACE model for propensity of ever being married

	Males	Females
c^2	.16	.22
h^2	.56**	.55**

family background endowments.

Shared-environment associations arise because individuals who grow up in the same household are subject to similar processes of socialization, socioeconomic conditions of parents, and similar family environments, all of which are thought to affect important demographic, socioeconomic and psychological outcomes. Non-shared environmental associations reflect the individual-specific conditions that affect these outcomes, while genetic associations originate through influences of inherited traits and predispositions on behavior and traits. The standard behavioral genetic model assumes that many genes contribute to a phenotype. This model is often denoted as an ACE model since it additively accounts for genetic (A), common environmental (C), and non-shared environment/error sources (E) of variance. The decomposition resulting from this model then factors the observed within-population variance into a genetic component (heritability, h^2) and a shared environmental component (c^2). The former measures the proportion of total phenotypic variance related to (additive) genetic variation across individuals, and the latter reflects the proportion of the total variance related to differences in shared-environmental conditions between twin pairs.

In this context, we estimate a standard additive behavioral genetic model on the propensity of ever being married. We control in the model for age. Table 3 shows the estimates for males and females.

The results are remarkably consistent for men and women and show that more than half of the variation in the heritability of ever being married can be explained by genetic influences. The influence of shared environmental effects is much less and varies between 16% for men and 22% for women.

5 Summary and Discussion

The model outlined above takes a different route than many other existing mortality models that include individual unobserved heterogeneity in general, as well as models that investigate whether the effect of marriage on mortality can be traced back to selective versus protective mechanisms. In particular, the fixed-effect approach in our analyses offers several interesting possibilities and new interpretations that are not possible with earlier applications.

First, the consistent estimation of the parameters of interest, β, which measure the impact of individual characteristics on the level of mortality, does not require the assumption of independence between the distribution of unobserved frailty z_j and the individual characteristics in x_{ij}. Furthermore, our analysis does not rely on specific distributional assumptions about unobserved heterogeneity. Our model is therefore particularly appropriate when the observed characteristics x_{ij} contains the outcome of individual decisions, such as marriage. Many of these characteristics will be correlated with unobserved individual characteristics such as health, earnings ability, etc.

Second, the estimation in Eq. (11) reveals two distinct effects of how the level of mortality is influenced by socioeconomic determinants: (a) the coefficient β directly reveal the dependence of the mortality hazard at some age a on the contemporaneous socioeconomic characteristics at age a; (b) the coefficients γ_0, γ_1,... in Eq. (11) reflect the effect of past socioeconomic conditions, or the socioeconomic history, of a respondent on mortality. This second effect is present if individual-specific heterogeneity exists in addition to twin-pair specific heterogeneity and different mortality levels in the past can lead to a differential selective pressure within twin pairs.

The combination of these two effects allows the evaluation of two different impacts of marriage on mortality. On one hand, 'being married' is associated with a *lower* mortality risk at any age because marriage may have protective (or other mortality reducing) effects. On the other hand, having been married for a long time in the past will tend to *increase* the mortality risk at age a among the survivors up to this age because the lower mortality in the past has lead to a weaker selection pressure towards low frailty in the past. Alternatively, if the fact of 'having been married in the past' is a causal determinant of individual-specific heterogeneity (like marriage 'produces' less frail individuals over time, for instance, due to the cumulative effect of better nutrition or social companionship), the impact of past marriage histories will be reverse to the above selection pattern. Because our estimation does not *a priori* impose either 'selection' or 'health protection' through past marriage, our model is able to identify these two mechanisms through their differential effect associated with past marriage on the conditional probability in Eq. (11).

Our estimations show that the fixed-effect approach and the standard survival models with and without incorporating random effects yield very different results on how mortality may be correlated with marriage. The piecewise-constant and the Gompertz estimation with and without unobserved heterogeneity confirm, not surprisingly, with the well-known pattern of mortality by marital status: for twins, similar to the general population, married men and women have the lowest risk of death. As we have indicated earlier, these standard estimations of the relationship between marital status and mortality are hampered, on one side, by unobserved differences among individuals that affect both the chance on the marriage market as well as for example their health status, and on the other hand, by the strong distributional assumptions about how these unobserved characteristics are related to the right-hand explanatory variables in the models. Moreover, the estimates obtained from the fixed-effect approach and the fact that our model does not yield similar results, indicate that these assumption do not necessarily hold and may bias the results.

Our results indicate that the fixed-effect approach is of a great advantage in the case when we can fully control for the effect of important unobserved characteristics such as in the case of MZ twins who share not only the same socioeconomic background, but share also the same genetic endowment which affects their chances on the marriage market as well as their survival (see for example Behrman et al. 1994 who write about genetic endowments as determinants of the marriage market). For DZ twins, for whom this 'full' control of important unobserved characteristics is not present, the fixed-effect survival models yields more or less similar results to the standard approaches.

Finally, our results based on the fixed-effect approach and data for MZ twins reveal an impor-

tant sex difference on mortality differentials by marital status. Mortality differentials by marital status diminish and are not any more statistically significant for MZ male twins, while for female MZ twins the model estimates that divorced and widowed women have lower risk of death than married women. In our opinion, these results indicate that the beneficial effect of marriage on mortality found in many conventional studies is primarily due to selection rather than to protective mechanisms. In particular, if it were marital protection, we would not expect to find differences between the fixed-effect model applied to MZ and DZ twins. The differences founded in Table 2, however, suggest that selection in the marital market is correlated with mortality relevant traits, and specifically traits that have a genetic origin.

The result for women also support this conclusion. According to the fixed-effect model estimated for MZ twins, divorced and widowed women have lower risk of death than marriage women (the difference is statistically significant), while the difference between single women and the reference group is not statistically significant. This pattern can be explained by the fact that women may select spouses on the basis of characteristics that are unobservable for us, but which are highly correlated with the status of these men in the society, or are for example, highly valued by employers (see for example Ginther and Zavodny 2001). Moreover, highly educated men and higher wage men are more likely to get married (Ginther and Zavodny 2001; Koskinen and Martelin 1994). It is plausible that assortative mating may contribute to the marriage premium if men select those women as spouses who actually increase their own productivity. When the husband dies or the woman divorces, there is an allocation of resources from which the widowed or divorced women may benefit. Moreover, there is evidence that the increased income in a union accounts much more for the protective effect of marriage for women rather than for men.

A Individual-specific heterogeneity

Potentially different histories of the twin 1 and twin 2 within pair j are particularly relevant if we further extend our model and additionally allow for individual-specific heterogeneity in addition to the twin-pair specific heterogeneity represented by the term z_j. In particular, once individual-specific heterogeneity is introduced, a different history $X_{ij}(a)$ between twins implies that at any age a the two twins will be subject to differential selection through mortality. In particular, twins who have been subject to higher mortality in the past will, on average, tend to be selected stronger towards low frailty individuals than twins who were subject to a less severe mortality regime in the past.

It turns out that under certain assumptions this differential selection can be identified with our fixed-effect estimation. In particular, this identification is possible by combining the contemporaneous characteristics at the age at first death, $x_{ij}(A_j^F)$, with information on the history of mortality determinants $X_{ij}(A_j^F)$ for each twin within a twin pair up to the age at first death A_j^F within twin pair j. In order to see this, assume an extension of our initial model in Eq. (6) that contains time-varying individual characteristics $x_{ij}(a)$ and additionally an individual-specific heterogeneity term e_{ij} as

$$\mu_{ij}(a) = h(a) \cdot z_j \cdot e_{ij} \cdot \exp(x_{ij}(a)\beta). \tag{9}$$

In lieu of any feasible alternative to sweep out the individual-specific heterogeneity we combine fixed and random effect estimation. The former sweeps out the twin-pair specific unobserved heterogeneity z_j, and the latter incorporates the unobserved individual-specific heterogeneity e_{ij}. In particular, we assume that e_{ij} is independent of x_{ij}, z_j and that is has an expectation of one at age zero, that is, we assume that $E[e_{ij}|X_{ij}, z_j, a = 0] = 1$. Due to selective mortality over time, this expectation will be different from one at any age $a > 0$, and the extent of this difference will depend on the level of mortality (Vaupel et al. 1979; Vaupel and Yashin 1985). In general, therefore, mortality over time will exert a selection towards less frail individuals and $E[e_{ij}|X_{ij}, z_j, a]$ will be less than one, and the difference will be larger for the twin within pair j who faced higher levels of mortality in the past.

Common approaches to include this selective force of mortality in the analyses has been to assume a specific distribution of the unobserved frailty e_{ij}, as for instance the gamma-distributed relative frailty models discussed above. In this paper, we take an alternative approach. We assume that the conditional expectation $E[e_{ij}|X_{ij}, z_j, a]$ can be decomposed—at least to a first approximation—into two terms, one depending only on age a and the twin-pair specific frailty z_j and the other depending only on age a and the history of socioeconomic determinants X_{ij}, as

$$\log E[e_{ij}|X_{ij}, z_j, a] = g(a, z_j) + k(a, X_{ij}). \tag{10}$$

For instance, if e_{ij} is gamma-distributed, then the decomposition in Eq. (10) is obtained as a first order approximation.[5]

The conditional probability $\Pr(D_j^{F1} = 1|x_{1j}, x_{2j}, A_j^F)$ that twin 1 is the first to die within pair j conditional on the age at first death A_j^F in this extended model with individual-specific and twin-pair specific heterogeneity then follows as

$$\begin{aligned}\Pr(D_j^{F1} = 1|X_{1j}, X_{2j}, A_j^F) &= \frac{E[\mu_{1j}|X_{1j}, z_j, A_j^F]}{E[\mu_{1j}|X_{1j}, z_j, A_j^F] + E[\mu_{2j}|X_{2j}, z_j, A_j^F]} \\ &= \frac{1}{1 + \exp[-\Delta x_j(A_j^F)\beta - \Delta k_j(A_j^F)]},\end{aligned}$$

where $\Delta k_j(A_j^F) = k(A_j^F, X_{1j}) - k(A_j^F, X_{2j})$. The above probability again represents a logistic regression. The only modification to our earlier analysis is that the logistic regression not only includes the within-pair difference in the individual characteristics at the age at first death, $\Delta x_j(A_j^F)$, but additionally also the difference in the histories of socioeconomic mortality determinants $\Delta k_j(A_j^F)$ that are modulated through the function $k(.)$.

[5]To see this, consider a simple gamma-distributed relative-frailty model with fixed covariates x_{ij}. The analyses in Vaupel et al. (1979) then imply that $E[e_{ij}|x_{ij}, z_j, a] = [1 + \sigma^2 z_j \exp(x_{ij}\beta) \int_0^a h(\alpha)d\alpha]^{-1}$. A first-order Taylor approximation to the logarithm of this expectation around $z_j = 1$ and $x_{ij} = 0$ yields

$$\log E[e_{ij}|x_{ij}, z_j, a] \approx \log(1 + \sigma^2 H(a)) - \frac{\sigma^2 H(a)}{1 + \sigma^2 H(a)} z_j - \frac{\sigma^2 H(a)}{1 + \sigma^2 H(a)} x_{ij}\beta,$$

where $H(a) = \int_0^a h(\alpha)d\alpha$. The expectation $\log E[e_{ij}|x_{ij}, z_j, a]$ can thus be decomposed to a first approximation as required in Eq. (10) into a term that depends only on $H(a)$ and unobserved frailty z, and a second term that depends only on $H(a)$ and the behavioral mortality determinants in x_{ij}.

In many cases, the unknown function $k(.)$ can be approximated by a polynomial function. The parameters in the above model can then be estimated via a logistic regression

$$\text{logit}(D_j^{F1} = 1|X_{1j}, X_{2j}, A_j^F) = \Delta x_j\beta + \Delta X_j\gamma_0 + \Delta X_j A_j^F\gamma_1 + ... \tag{11}$$

where we have used the approximation $\Delta k_j(A_j^F) = k(A_j^F, X_{1j}) - k(A_j^F, X_{2j}) \approx \Delta X_j\gamma_0 + \Delta X_j A_j^F\gamma_1 +$ .... The term $\Delta X_j = X_{1j} - X_{2j}$ denotes the difference between the socioeconomic histories and $\Delta x_j = x_{1j} - x_{2j}$ denotes the difference in the contemporaneous individual characteristics, and both differences are evaluated the age of first death A_j^F within the twin pair.

References

Andersen, E. B. (1970). Asymptotic properties of conditional maximum likelihood estimation. *Journal of the Royal Statistical Society, Series B* 32: 281–301.

Ashenfelter, O. and Krueger, A. (1994). Estimates of the economic return to schooling from a new sample of twins. *American Economic Review* 84(5): 1157–1173.

Ashenfelter, O. and Rouse, C. (1998). Income, schooling and ability: Evidence from a new sample of identical twins. *Quarterly Journal of Economics* 113(1): 153–284.

Behrman, J. R., Kohler, H.-P., Jensen, V., Pedersen, D., Petersen, I., Bingley, P. and Christensen, K. (2011). Does more schooling reduce hospitalization and delay mortality? New evidence based on Danish twins. *Demography* doi:10.1007/s13524-011-0052-1. Published online August 13, 2011.

Behrman, J. R. and Rosenzweig, M. R. (1999). Ability biases in schooling returns and twins: A test and new estimates. *Economics of Education Review* 18(2): 159–167.

Behrman, J. R. and Rosenzweig, M. R. (2002). Does increasing women's schooling raise the schooling of the next generation? *American Economic Review* 92(1): 323–334.

Behrman, J. R. and Rosenzweig, M. R. (2004). Returns to birthweight. *Review of Economics and Statistics* 86(2): 586–601.

Behrman, J. R., Rosenzweig, M. R. and Taubman, P. (1994). Endowments and the allocation of schooling in the family and in the marriage market: The twins experiment. *Journal of Political Economy* 102(6): 1131–1173.

Behrman, J. R., Rosenzweig, M. R. and Taubman, P. (1996). College choice and wages: Estimates using data on female twins. *Review of Economics and Statistics* 73(4): 672–685.

Behrman, J. R. and Taubman, P. (1976). Intergenerational transmission of income and wealth. *American Economic Review* 66(2): 436–440.

Burgoa, M., Regidor, E., Rodriguez, C. and Gutierrez-Fisac, J. L. (1998). Mortality by cause of death and marital status in Spain. *European journal of public health* (8): 37–42.

Christensen, K. (2001). Biological materials in household surveys: The inferface between epidemiology and genetics. In: Finch, C. E., Vaupel, J. W. and Kinsella, K. (eds.). *Cells and Surveys: Should Biological Measures Be Included in Social Science Research*. Washington, D.C.: The National Academies Press, 42–63.

Conley, D., Strully, K. and Bennett, N. (2003). A pound of flesh or just proxy? Using twin differences to estimate the effects of birth weight on (literal) life chances. NBER Working Paper No. 9901.

Durkheim, E. (1951). *Suicide: A Stidy in Sociology*. New York: The Free Press.

Farr, W. (1858). The influence of marriage on the mortality of the french people. In: Hastings, G. W. (ed.). *Transactions of the National Association for the Promotion of Social Science*. London: John W. Parker, 504–513.

Ginther, D. K. and Zavodny, M. (2001). Is the male marriage premium due to selection? the effect of shotgun weddings on the return to marriage. *Journal of Population Economics* 14(2): 313–328.

Goldman, N. (1993). Marriage selection and mortality patterns: Inferences and fallacies. *Demography* 30(2): 189–208.

Gourieroux, C. and Monfort, C. (1995). *Statistics and Econometric Models*, vol. 1 & 2. Cambridge: Cambridge University Press.

Gove, W. R. (1973). Sex, marital status, and mortality. *American Journal of Sociology* 79(1): 45–66.

Hauge, M. (1981). The Danish twin register. In: Mednich, S. A., Baert, A. E. and Bauchmann, B. P. (eds.). *Prospective Longitudinal Research: An Empirical Basis for the Primary Prevention of Psychological Disorders*. Oxford: Oxford University Press, 217–221.

Hauge, M., Harvald, B., Fischer, M., Gotlieb-Jensen, K., Juel-Nielsen, N., Raebild, I., Shapiro, R. and Videbech, T. (1968). The Danish twin registers. *Acta Geneticae Medicae et Gemellologiae* 2: 315–331.

Horwitz, A. V., White, H. R. and Howell-White, S. (1996). Becoming married and mental health: A longitudinal study of a cohort of young adults. *Journal of Marriage and the Family* 58: 895–907.

House, J. S., Landis, K. R. and Umberson, D. (1988). Social relationships and health. *Science* 241(4865): 540–545. doi:10.1126/science.3399889.

Hsiao, C. (1986). *Analysis of Panel Data*. Cambridge: Cambridge University Press.

Hu, Y. and Goldman, N. (1990). Mortality differentials by marital status: An international comparison. *Demography* 27(2): 233–250.

Jacobs, J. A. and Gerson, K. (2004). *The Time Divide: Work, Family and Gender Inequality*. Cambridge: Harvard University Press.

Kisker, E. E. and Goldman, N. (1987). Perils of single life and benefits of marriage. *Social Biology* 34: 135–152.

Kohler, H.-P., Behrman, J. R. and Schnittker, J. (2011). Social science methods for twins data: Integrating causality, endowments and heritability. *Journal of Biodemography and Social Biology* 57(1): 88–141. doi:10.1080/19485565.2011.580619.

Kohler, H.-P. and Rodgers, J. L. (2003). Education, fertility and heritability: Explaining a paradox. In: Wachter, K. W. and Bulatao, R. A. (eds.). *Offspring: Human Fertility Behavior in Biodemographic Perspective*. Washington, D.C.: The National Academies Press, 46–90. `http://books.google.com/books?id=poXXt7ta73cC&pg=PA46`.

Kohler, H.-P., Rodgers, J. L. and Christensen, K. (1999). Is fertility behavior in our genes: Findings from a Danish twin study. *Population and Development Review* 25(2): 253–288. `http://www.jstor.org/stable/172425`.

Koskinen, S. and Martelin, T. (1994). Wha are socioeconomic mortality differences smaller among women than among men? *Social Science and Medicine* 38(10): 1385–1396.

Kravdal, O. (2001). The impact of marital status on cancer survival. *Social Science and Medicine* 52(3): 357–368.

Kyvik, K. O., Christensen, K., Skytthe, A., Harvald, B. and Holm, N. V. (1996). The Danish twin register. *Danish Medical Bulletin* 43(5): 465–470.

Kyvik, K. O., Green, A. and Beck-Nielsen, H. (1995). The new Danish twin register: Establishment and analysis of twinning rates. *International Journal of Epidemiology* 24: 589–596.

Lillard, L. A. (1993). Simultaneous equations for hazards: marriage duration and fertility timing. *Journal of Econometrics* 56: 189–217.

Lillard, L. A. and Panis, C. W. A. (1996). Marital status and mortality: The role of health. *Demography* 33(3): 313–327.

McGue, M. and Christensen, K. (1997). Genetic and environmental contributions to depression symptomatology: Evidence from Danish twins 75 years of age and older. *Journal of Abnormal Psychology* 106(3): 439–448.

McGue, M. and Christensen, K. (2001). The heritability of cognitive functioning in very old adults: Evidence from Danish twins aged 75 and older. *Psychology and Aging* 16: 272–280.

Morgan, M. (1980). Marital status, health, illness and service use. *Social Science and Medicine* 14A: 633–643.

Murray, J. E. (2000). Marital protection and marital selection: Evidence from a historical-prospective sample of american men. *Demography* 37(4): 511–521.

Parker-Pope, T. (2010). Is marriage good for your health? *New York Times Magazine*, April 12, 2010, `http://www.nytimes.com/2010/04/18/magazine/18marriage-t.html`.

Pienta, A. M., Hayward, M. D. and Jenkins, K. R. (2000). Health consequences of marriage for the retirement years. *Journal of Family Issues* 21(5): 559–586.

Preston, S. H. and Taubman, P. (1994). Socioeconomic differences in adult mortality and health status. In: Martin, L. and Preston, S. H. E. (eds.). *Demography of Aging*. Washington, D.C.: National Academy Press, 279–318.

Rendall, M., Weden, M., Favreault, M. and Waldron, H. (2011). The protective effect of marriage for survival: A review and update. *Demography* 48: 481–506. doi:10.1007/s13524-011-0032-5.

Rutledge, T., Matthews, K., Lui, L.-Y., Katie, S. and Cauley, J. (2003). Social networks and marital status predict mortality in older women: Prospective evidence from the study of osteoporotic fractures (SOF). *Psychosomatic Medicine* 65: 688–694.

Simon, R. (2002). Revisiting the relationships among gender, marital status, and mental health. *American Journal of Sociology* 107(4): 1065–1096.

Trovato, F. and Lauris, G. (1989). Marital status and mortality in Canada: 1951-1981. *Journal of Marriage and the Family* 51: 907–922.

Tucker, J. S. and Anders, S. (2001). Social control of health behaviors in marriage. *Journal of Apllied Social Psychology* 31(3): 467–485.

Umberson, D. (1987). Family status and health behaviors: Social control as a dimension of social integration. *Journal of Health and Social Behavior* 28: 306–319.

Vaupel, J. W., Manton, K. G. and Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* 16(3): 439–454.

Vaupel, J. W. and Yashin, A. I. (1985). Heterogeneity's ruses: Some surprising effects of selection on population dynamics. *American Statistician* 39: 176–185.

Waite, L. J. (1995). Does marriage matter? *Demography* 32(4): 483–507.

Waite, L. J. and Gallagher, M. (2000). *The Case for Marriage: Why Married People are Happier, Healthier, and Better off Financially*. New York: Doubleday.

Waite, L. J. and Lehrer, E. L. (2003). The benefits from marriage and religion in the United States: A comparative analysis. *Population and Development Review* 29(2): 255–275.

Welon, Z., Bielicki, T., Rogucka, E. and Malina, R. M. (1999). Effect of education and marital status on premature mortality among urban adults in Poland, 1988-1989. *American Journal of Human Biology* 11(3): 397–403.

Yashin, A. I. and Iachine, I. A. (1995). Genetic analysis of duration: correlated frailty model applied to survival of Danish twins. *Genetic Epidemiology* 12: 529–38.

Zick, C. D. and Smith, K. R. (1991). Marital transition, poverty, and gender differences in mortality. *Journal of Marriage and the Family* 53(2): 327–336.