I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11448

A Unique Bond: Twin Bereavement and Lifespan Associations of Identical and Fraternal Twins

Gerard J. van den Berg
Bettina Drepper

APRIL 2018

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11448

A Unique Bond: Twin Bereavement and Lifespan Associations of Identical and Fraternal Twins

Gerard J. van den Berg
University of Bristol, IFAU-Uppsala, IZA, ZEW and CEPR

Bettina Drepper
Tilburg University and IZA

APRIL 2018

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

A Unique Bond: Twin Bereavement and Lifespan Associations of Identical and Fraternal Twins*

Empirical analyses of twin mortality often use models with dependent unobserved frailty terms capturing genetic and childhood environmental determinants. This ignores that mortality rates can be co-dependent due to bereavement effects, i.e. to a time-dependent causal effect of the loss of the co-twin on the mortality rate of the surviving twin. We develop a novel methodology and perform an empirical analysis based on a comprehensive model incorporating both types of dependence. We prove full identification without functional-form restrictions and we estimate models with data on twin pairs from the Danish Twin Registry. Among men, the loss of an identical co-twin at age 75 causally reduces the remaining lifetime on average by more than a year. This bereavement effect is less severe among non-identical twins or if the loss occurs at a higher age. Estimates of correlations between the frailty terms by zygosity and the ensuing implications for the relative importance of mortality determinants are highly sensitive to whether bereavement is taken into account.

JEL Classification: C41, C32, I14

Keywords: mortality, longevity, duration, frailty, genetic determinants, hazard rate, identification, loss of co-twin

Corresponding author:
Gerard J. van den Berg
School of Economics, Finance and Management
University of Bristol
2C3, The Priory Road Complex
Priory Road, Clifton
Bristol, BS8 1TU
United Kingdom

E-mail: gerard.vandenberg@bristol.ac.uk

* We acknowledge valuable comments and suggestions by Bo Honoré, Petyo Bonev, Georgios Effraimidis and Lena Janys. The Danish Twin Registry and Kaare Christensen kindly allowed us to use their data. Thanks to the Humboldt Foundation for financial support through their Alexander von Humboldt Professorship Prize program. The research of Bettina Drepper is financially supported by the Netherlands Organisation for Scientific Research (NWO) through Vici grant 453-11-002. Previous title: "A Unique Bond: Does Losing Your Co-Twin Affect Your Remaining Life-Span?"

1 Introduction

Twin mortality has been an important research topic since many decades. First and foremost, the comparison of the association of lifetimes of identical twins with the association of lifetimes of fraternal twins is informative on the role of genetic mortality determinants. This extends to cause-specific mortality and rates at which illnesses occur. In the literature, the most elaborate approach involves the estimation of bivariate survival models with unobserved mortality determinants called "frailty terms". The latter terms are individual-specific but may be stochastically related among twins, and the correlation is allowed to depend on the zygosity of the twin pair (e.g., see Hougaard et al., 1992a,b; Yashin and Iachine, 1995a; Wienke et al., 2001). Frailty terms allow for a convenient interpretation as a summary measure of the effects of the underlying genetic predisposition and childhood environmental determinants. In bivariate models with frailty terms, the association of lifetimes of twins is driven by the correlation of their frailty terms. If frailty terms are more strongly correlated among identical twins then this indicates that genetic background is important.

However, there is a second major reason for why twin lifetimes can be correlated, and this concerns causal bereavement effects of the loss of the co-twin on the mortality rate of the surviving twin. Due to the unique bond that twins share, the loss of the co-twin can have severe effects on the surviving twin. Studies of bereaved twins document that the loss of the co-twin causes psychological stress that can also lead to a health deterioration.[1] Ultimately this may lead to a higher mortality rate. Of course one may think of other events in the adult life of twins that affect the association of their lifetimes, but bereavements are particularly relevant for the analysis of mortality because by definition bereavements mostly occur in the age span where mortality rates are high, that is, at high ages. This is why bereavements are potentially important drivers of lifetime associations especially among the elderly. Ignoring bereavement effects in twin mortality analysis may lead to biased results. To see this, note that bereavement effects may be present even if

[1] Segal et al. (1995, 2002); Segal and Ream (1998) and Woodward (1988) document how the loss of the co-twin can cause severe emotional stress. Besides feelings of despair, depersonalization (numbness, shock), rumination (preoccupation with the deceased) and loss of control, bereaved twins also show symptoms such as loss of appetite and vigor and other physical symptoms (Segal and Blozis, 2002). Chronic stress due to bereavement is a major cause of disease; see e.g. Stroebe and Stroebe (1987); Stroebe et al. (1993), Sanders (1980; 1999) and Selye (1936; 1955).

genetic and early-life environmental characteristics are irrelevant. In that case, the association of twins' lifetimes may be wrongly attributed to those characteristics. What is more, in real life, bereavement effects may be stronger among identical twins because of how their social environment reacts to the loss, and this strengthens the association of their lifetimes, which in turn may lead to incorrect inference on the effects of genes on mortality.[2]

Adverse bereavement effects on mortality are of interest in their own right. They involve health care costs and they may inspire policymakers to design treatments to ameliorate their size. Twin bereavement effects on mortality have been analyzed by e.g. Tomassini et al. (2001; 2002) and Hougaard et al. (1992a). However, these and other existing empirical studies on this topic base their results on the estimation of univariate survival models for one member of each twin pair, where an indicator of the realized mortality of the co-twin is included as an exogenous time-dependent covariate. Clearly, this ignores longevity dependence due to correlated unobserved individual characteristics. This drawback is the mirror image of the above-mentioned limitation of the frailty approach, namely that one of the two types of dependence of lifetimes is ruled out by construction.[3] As already pointed out by Hougaard et al. (1992a), this may bias the results.

Given the interest in each type of dependence between twin lifetimes, there is a need for an integrative approach that provides consistent estimates for each type of dependence in the presence of the alternative type. Indeed, in her discussion of Hougaard et al.'s (1992a) paper, Flournoy (1992) argues that a super-model is needed that accounts for both effects simultaneously: the bereavement effect and the influence of unobserved correlated factors. This is a challege precisely because it is difficult to disentangle two reasons for one observed association.

This paper develops and applies such an integrative approach. We introduce a bivariate survival model for twin lifetimes which incorporates bereavement effects as well as

[2]One could claim that stronger bereavement effects among identical twins are to some extent due to genetic factors. However, this stretches the range of what constitute genetic effects and this would limit the usefulness of twin studies to assess the contribution of genetic factors in the population as a whole. We return to this in the final section of the paper.

[3]Tomassini et al. (2002; 2001) match each bereaved twin to two unbereaved twins based on zygosity, age, and sex and compare the two resulting mortality rates after the age when bereavement takes place. This ignores the endogeneity of the time of bereavement caused by shared genetic factors, or in other words, the dependence of the underlying frailty terms.

unobserved frailty terms that may capture genetic and childhood effects. As seen from a single twin, the model includes mortality of the co-twin as a time-dependent covariate. In addition, each of the twins' own mortality rates is allowed to depend on an individual-specific frailty term that may be correlated with the co-twin's frailty term. Note that this accounts for the endogeneity of the co-twin's mortality as a covariate. Each frailty term is assumed to affect the corresponding individual mortality rate in a multiplicative fashion. Thus, the model can be seen as an extension of the familiar Mixed Proportional Hazard (MPH) model (see e.g. Van den Berg, 2001).

In the paper we formally prove that with minimal covariate variation between twin pairs, all components of the model are identified from the observable joint distribution of twin lifespans conditional on observed covariates. This includes the identification of the bereavement effect and the joint distribution of the frailty terms. Identification does not require covariate values to differ by twin within the twin pair, and it does not require functional-form assumptions on the force of mortality or the distribution of the frailty terms. In general, in the estimation of models with unobserved confounders, identification is a valuable property, as it implies that the results are not driven by such functional-form assumptions. Our identification result extends to cause-specific mortality.

We estimate the model with data on 9,270 twin pairs in the Danish Twin Registry born in Denmark between 1873 and 1930. This includes 2,808 identical (monozygotic) and 6,462 fraternal (dizygotic) pairs. For our purposes, a major advantage of these cohorts is that right-censoring of lifetimes is relatively rare. Indeed, 81% of the twins are observed to die before 2004. In the data, the observed covariates do not vary within same-sex twin pairs.[4]

We also examine whether the estimated correlation of the twins' frailty terms is sensitive to whether bereavement is taken into account or not, for each zygosity. More in general, we show how estimated effects change when one of the two sources of lifetime dependence is ignored.

At least three strands of literature are connected to our study. First, as already mentioned, there is the literature on bivariate survival models with unobserved frailty terms, including applications to twin mortality. In the twin context, this adjusts standard

[4]The majority of the data comprises same-sex twin pairs, since collection of their records had priority in the early years of the Twin Registry which is the first ever nationwide twin register in the world.

variance-decomposition methods to deal with mortality outcomes. Secondly, there is the wider literature on bereavement effects on mortality. Most of this literature concerns conjugal bereavement (e.g., see Bowling, 1987; Lichtenstein et al., 1998; Lindeboom et al., 2002; Manor and Eisenbach, 2003; Van den Berg et al., 2011). These studies tend to find sizeable effects, reflecting both economic loss and emotional hardship. In contrast to spouses, most adult twins have separate families and separate economic support systems, so that twin bereavement effects can be expected to mostly reflect emotional hardship. This suggests that by comparing the magnitudes of conjugal and twin bereavement effects, we may identify the relative importance of the emotional component in the total conjugal effect. As such, our analysis makes a contribution to wider bereavement literature.

A third relevant body of work concerns the identification of bivariate duration models that contain causal effects as well as unobserved heterogeneity. In the so-called Timing-of-Events approach, the realization of some event (say, a treatment) may causally affect the hazard rate of some duration outcome variable. The duration until treatment has its own hazard rate, and the Timing-of-Event model postulates MPH-type specifications for each of the two hazard rates, where the unobserved determinants of the hazard rates may be mutually dependent. Clearly, this resembles our twin bereavement model. Abbring and Van den Berg (2003) prove identification without functional-form assumptions. However, this requires observed covariates to have different effects on each of the two hazard rates of the individual. In applications of the latter model, it generally does not make sense to expect covariate effects on treatment and outcome to be identical. In contrast, in our current setting, same-sex twins have identical covariate values. Consequently, their identification result cannot be straightforwardly extended to our setting. At the same time, in our setting, the two hazard rates naturally have symmetric specifications as functions of their determinants, whereas the Timing-of-Event model does not assume symmetry.[5]

The outline of the paper is as follows. In Section 2 we introduce the mortality model with bereavement effects and we prove identification. Section 3 presents the twin dataset from the Danish Twin Registry while Section 4 discusses the estimation method. Subsequently, the estimation results are in Section 5. Section 6 concludes.

[5]Interestingly, the Timing-of-Events approach has been fruitfully used to study conjugal bereavement effects on mortality, where there is no compelling reason to impose symmetry restrictions across the male and female mortality rates, and where covariate values differ within couples (see Van den Berg et al., 2011; Gourieroux and Lu, 2015; Lu, 2017).

2 Model and identification result

In this section we introduce a new bivariate model for twin life-spans. Each twin is exposed to the risk of dying at every age $t \in [0, \infty)$, given that he has reached that age. Since we are interested in measuring the causal effect of the end of one life-span on the subsequent residual life-span of the other (the bereavement effect), we specify the mortality rate (or "hazard") of each twin $j = 1, 2$ conditional on the realization of the life-span of the co-twin $T_k = t_k$. In addition, we condition on observable characteristics x of the twin pair and the realization of frailty terms V_j.

Model 1. *The hazard rates of* $T_1|(T_2 = t_2, x, V_1)$ *and* $T_2|(T_1 = t_1, x, V_2)$ *are given by*

$$\theta(t|T_2 = t_2, x, V_1) = \lambda(t)\phi(x)\delta(t, t_2, x)^{I(t>t_2)}V_1$$
$$\theta(t|T_1 = t_1, x, V_2) = \lambda(t)\phi(x)\delta(t, t_1, x)^{I(t>t_1)}V_2,$$

where the vector of frailties $V = (V_1, V_2)'$ *is assumed to be drawn from a bivariate distribution* $G(v_1, v_2)$ *and the bereavement effect function is multiplicative in two of its arguments* $\delta(t, t_k, x) = \delta_a(t - t_k)\delta_b(t_k, x)$.

The function $\lambda(t)$ captures the dependence of the mortality hazard on age and $\phi(x)$ incorporates the effect of covariates. $I(t > t_k)$ denotes the indicator function that is equal to one when the loss of the co-twin has occurred and zero otherwise. As long as both twins are alive, each twin j faces the mortality hazard $\lambda(t)\phi(x)V_j$. Once the co-twin has died, the mortality hazard of the surviving twin is rescaled by $\delta_a(t - t_k)\delta_b(t_k, x)$, reflecting the bereavement effect. Here, the first multiplicative term δ_a describes the dependence of the bereavement effect on the time passed since the loss occurred, while δ_b accounts for the dependence on the age at the time of bereavement and covariates x.[6]

Conditional on observed twin pair characteristics x, Model 1 allows for two types of dependencies between life-spans T_1 and T_2. The first is introduced through the joint distribution of V_1 and V_2 capturing similarities in genetic makeup and childhood experiences of the two twins. The second is introduced through the function $\delta(t, t_k, x)$ reflecting the

[6]The identification result below can be straightforwardly extended to cases where the bereavement effect function differs between the two durations. Thus, if the two spells can be distinguished in the data, it is possible to identify two separate bereavement effects $\delta_1(t, t_2, x)$ and $\delta_2(t, t_1, x)$.

effect of bereavement. Note that conditional on x and V, the only dependence between life-spans T_1 and T_2 comes from the bereavement effect function $\delta(t, t_k, x)$. Consequently, this function can be given a causal interpretation as the effect of the end of one life-span on the remaining length of the other.

The bereavement effect in Model 1 can be said to generate a local dependence between T_1 and T_2, as it only affects the hazard rate of the surviving twin after the loss has occurred. In contrast to this, the time-constant unobserved factors V give rise to a global dependence. This reflects the fact that genetic dispositions and characteristics shaped during childhood influence the mortality hazard of the two twins over their whole lifespan. The difference between global and local dependence is key to the identification of our model.

The local nature of the bereavement effect in Model 1 rules out anticipatory effects. In particularly, it rules out a scenario in which a twin anticipates the future date of death of his co-twin and is affected by this knowledge to the degree that his own mortality hazard today is affected. Furthermore, the unobservable influences V are assumed to be time-constant, thus ignoring twin-pair specific unobservable shocks, such as local epidemics or major events within the extended family.

Model 1 is a symmetric version of the Timing-of-Events model (Abbring and Van den Berg, 2003) in the sense that the functions λ, ϕ and δ are identical for the two twins. Moreover, the lifespan of each twin can potentially be affected by the death of its co-twin. We assume that the bereavement effect function is multiplicative in a function of the time $t-t_k$ since the bereavement. We do not impose that x varies within twin pairs. As we shall see, in the 1873-1930 birth cohorts in our data, only pair-specific covariates are observed, with the exception of the twin's sex which is the same within the majority of twin pairs. If the value of $\phi(x)$ is identical within each twin pair then the identification approach used for the Timing-of-Events model cannot be adopted.

As prerequisites for identification of Model 1, we impose the following regularity assumptions:

Assumption 1. *The vector x is k-dimensional with $1 \leq k < \infty$ and $\phi : \mathcal{X} \to U \subset (0, \infty)$. The set $\mathcal{X} \subset \mathbb{R}^k$ contains at least two values.*

Assumption 2. $\delta_a : \mathbb{R}_+ \to (0, \infty)$ *with* $\lim_{s \downarrow 0} \delta_a(s) < \infty$ *and for* $\delta_b : [0, \infty) \times \mathcal{X} \to (0, \infty)$

it holds that $\nexists\ c \in (0,\infty)$ *s.t.* $\delta_b(0,x) = c\phi(x)^{-1}\ \forall x \in \mathcal{X}$.

Assumption 3. *For the function* $\lambda : [0,\infty) \to (0,\infty)$ *it holds that for all* $t \in (0,\infty)$ $\lim_{s\downarrow t} \lambda(s) < \infty$ *and has integral* $\Lambda(t) := \int_0^t \lambda(\tau)\, d\tau < \infty,\ \forall\ t \geq 0$ *and further*

$$\tilde{\Lambda}(t,s) := \int_s^t \lambda(\tau)\delta_a(\tau - s)\, d\tau < \infty, \quad \forall\ \{(t,s) \in [0,\infty)^2 : t > s\}.$$

For some a priori chosen t_0, t_0^* *and* x_0, *it holds that* $\int_0^{t_0} \lambda(\tau)\, d\tau = 1$, $\int_0^{t_0^*} \lambda(\tau)\delta_a(\tau)\, d\tau = 1$ *and* $\phi(x_0) = 1$.

Assumption 4. V *is an* $\mathbb{R}_+^2$*-valued time-invariant random vector* $(V_1, V_2)'$ *and is drawn from a distribution* G *which does not depend on* x *and has a finite positive mean.* G *is such that* $P(V \in (0,\infty)^2) = 1$. *Furthermore, for all* $(t,x) \in (0,\infty) \times \mathcal{X}$ $\lim_{s\downarrow t} E(V_j | T_j \geq s, T_k = t, x) = E(V_j | T_j \geq t, T_k = t, x)$.

Assumption 5. $\exists$ *an open set* $\Psi \in (0,\infty)^2$ *with* $t_1 > t_2\ \forall\ (t_1,t_2) \in \Psi$ *s.t. at all points* $(t_1,t_2) \in \Psi$ *the function* $\Delta(t_1,t_2,x) = \tilde{\Lambda}(t_1,t_2)\delta_b(t_2,x)$ *is continuously differentiable with respect to* t_2.[7]

For Assumption 1 a single dummy variable x suffices that does not need to vary across the two hazards, provided that it has an effect. In such a case, $\phi(x)$ takes on only two values on $\mathcal{X}$. Assumption 3 restricts the baseline hazard function to be continuous from the right for all $t \in (0,\infty)$. Note that this does not rule out the piecewise constant case or most functional forms. Furthermore, given that this property only has to hold for strictly positive values, functional forms with $\lim_{s\downarrow 0} \lambda(s) = \infty$ such as the Weibull function are not ruled out. However, the magnitude of the instantaneous bereavement effect must have a finite limit. Consequently, functional forms of δ_a with $\lim_{s\downarrow 0} \delta_a(s) = \infty$ are excluded in Assumption 2.

Proposition 1. *If Assumptions 1-4 are satisfied, then the functions* $\lambda, \phi, \delta_a, \delta_b$ *from Model 1 are non-parametrically identified from the distribution of* $(T_1, T_2)|x$.

The proof is in Appendix A.1 (for λ and ϕ) and Appendix A.2 (for δ_a and δ_b). Note that G remains undetermined in Proposition 1. This leads to:

[7] Alternative assumption 5: The open set $\Psi \in (0,\infty)^2$ could also exist for $t_1 < t_2\ \forall\ (t_1,t_2) \in \Psi$ s.t. at all points $(t_1,t_2) \in \Psi$ the function $\Delta(t_2,t_1,x)$ is continuously differentiable with respect to t_1.

Proposition 2. *If Assumptions 1-5 are satisfied, then Model 1, which is characterized by the functions* $G, \lambda, \phi, \delta_a, \delta_b$*, is non-parametrically identified from the distribution of* $(T_1, T_2)|x$*.*

The proof is in Appendix A.3.

In the empirical analysis with models that contain unobserved confounders, non-parametric identification is a valuable property, as it implies that the results are not fully driven by ad hoc functional-form assumptions. If the model is not non-parametrically identified then the estimation of a model with parametric functions $G, \lambda, \phi, \delta_a, \delta_b$ may give a priori sensible point estimates but these would be fully driven by the parametric functional forms. In particular, estimation with a different set of parametric functions may lead to identical point estimates.

We make two minor comments about the identification results. First, the results can be applied to the study of cause-specific mortality. The hazard rates then represent rates of mortality due to a specific cause. This requires that death due to other causes can be regarded as independent right-censoring of the duration until death to that particular cause. Secondly, the proofs of the propositions do not use the assumption that the marginal distributions of the frailty terms V_1 and V_2 are identical. Hence, this assumption is superfluous. For cases where there is some natural ordering of twins within twin pairs, this motivates a model extension in which the marginal distributions are allowed to be different. Alternatively, this overidentifying restriction may be used for a model specification test. It may be an interesting topic for further research to see if the overidentifying information can be used to identify models that are less restrictive insome other direction.

3 The Danish Twin Registry

The Danish Twin Registry was first established in 1954 with the goal of following up on all same-sex twins who were born since 1873 and who survived as twins at least until the age of 6. However, there is some selectivity in the very early cohorts, with twins who died young less likely to be included in the sample. Furthermore, covariates are barely observed if the pair did not survive as twins until January 1, 1943. Therefore, we restrict attention to twin pairs still alive at that date. We use cohorts from 1873 to 1930, assuring

that we observe uncensored lifespans for most twins before January 1, 2004, when our window of observation ends. While the registry contains some different-sex twin pairs, most effort was devoted to following up on same-sex and particularly monozygotic twin pairs. We refer to Skytthe et al. (2002) and Hauge et al. (1968) for detailed descriptions of the registry and the way in which it has been collected.

The resulting sample includes 2,870 monozygotic and 6,625 dizygotic twin pairs, 1,239 of which are different sex twin pairs. Twins still alive on January 1, 2004 or who emigrated at a prior date have right-censored lifespans. Overall, the death date is observed for 80% of the individuals in our sample. For each twin pair in our sample, we observe zygosity, sex, region of birth and date of birth. The information on zygosity has been shown to be highly accurate, with a misclassification rate below 5% (see Holm, 1983; Lykken, 1978). We restrict attention to an indicator for being born in Copenhagen to distinguish between rural and urban areas in Denmark. Additional distinctions between small towns and rural areas outside of Copenhagen proved to be uninformative.

In the Registry, each twin pair's two members are distinguished by labels 1 and 2. Since the data were manually copied from parish books, and since traditionally the first-born twin member was recorded first, it seems likely that twin 1 is the first-born and twin 2 the second-born. However, there is no solid evidence for this and therefore this variable is typically not used in studies with the Registry (e.g., see Hougaard et al., 1992b; Herskind et al., 1996). In our sample, a Kolmogorov-Smirnov test for equality of distribution functions of T_1 and T_2 fails to reject the null hypothesis of equal distributions in our sample (significance level 0.05). Therefore, we do not use this variable in our analyses.

For our purposes it is important to point out that Denmark did not witness major epidemics between 1873 and 2004. Cross-national comparisons reveal that Denmark stands out as the country with the lowest excess mortality for the 1918/1919 worldwide influenza pandemic (see Canudas-Romo and Erlangsen, 2008; Ansart et al., 2009). Furthermore, Denmark remained neutral in both World Wars, and despite being occupied by Germany during the Second World War, casualties were negligible compared to the rest of Europe.

4 Empirical implementation

In this subsection we explain how we estimate the model of Section 2 with the data of the Twin Registry. We choose flexible specifications for the model determinants λ, ϕ, δ and G and we estimate the various model versions with Maximum Likelihood.

The vector of frailties (V_1, V_2) for each twin pair is assumed to be drawn from a Cherian bivariate Gamma distribution. This family of distributions is often used for twin frailty terms in mortality models (e.g., see Yashin and Iachine, 1995b; Wienke et al., 2001, 2002) as it allows for the interpretation of the individual frailty term as the sum of a shared twin pair-specific term $\tilde{V}_0$ and an individual-specific term $\tilde{V}_j$:

$$V_j = \tilde{V}_0 + \tilde{V}_j \quad \text{for } j \in 1, 2.$$

Here, each term $\tilde{V}_1, \tilde{V}_2$ and $\tilde{V}_0$ is independently drawn from a Gamma distribution. With this structure, the bivariate Gamma distribution of (V_1, V_2) has identical marginal distributions, which makes sense. Their mean is normalized to one, and consequently the joint distribution of (V_1, V_2) can be fully described by two parameters: the variance σ^2 of V_j and correlation ρ of V_1 and V_2. The latter is computed as the ratio of the shared and total variation $\rho = \text{Var}(\tilde{V}_0)/(\text{Var}(\tilde{V}_0 + \tilde{V}_j)$. Recall that our sample includes monozygotic (MZ) and dizygotic (DZ) twin pairs. Accordingly, we estimate separate parameters for both types of zygosity: $\sigma^2_{MZ},\ \rho_{MZ}$ and $\sigma^2_{DZ},\ \rho_{DZ}$.[8]

For a twin pair with right-censored lifespans at t_1 and t_2 the bivariate survival function $S(t_1, t_2|x) = P(T_1 > t_1, T_2 > t_2|x)$ can now be expressed as

$$S(t_1, t_2|x) = \begin{cases} S^*(t_1, t_1|x) - \int_{t_2}^{t_1} S_{t_2}(t_1, \tau|x)\, d\tau & \text{, for } t_1 \geq t_2 \\ S^*(t_2, t_2|x) - \int_{t_1}^{t_2} S_{t_1}(\tau, t_2|x)\, d\tau & \text{, for } t_1 < t_2 \end{cases}$$

[8] A key assumption for the identification result in Section 2 is that G does not depend on covariates x. Zygosity will be excluded from x such that Propositions 1 and 2 can be applied to the separate samples of monozygotic and dizygotic twin pairs. This way, the two distributions G_{MZ} and G_{DZ} are identified.

$$\text{with} \quad S^*(t_1,t_2|x) = (1+\sigma^2\phi(x)[\Lambda(t_1)+\Lambda(t_2)])^{-\frac{\rho}{\sigma^2}} \\ (1+\sigma^2\phi(x)\Lambda(t_1))^{-\frac{(1-\rho)}{\sigma^2}}(1+\sigma^2\phi(x)\Lambda(t_2))^{-\frac{(1-\rho)}{\sigma^2}} \tag{1}$$

and with partial derivatives $S_{t_j}(t_1,t_2|x) = \frac{\partial S(t_1,t_2|x)}{\partial t_j}$ for $(j=1,2)$. For a few of the twin pairs the life-span of one twin is right-censored while the co-twin is observed to live past this censoring point. Here, right-censoring may occur e.g. due to emigration. As a consequence, bereavement could have occurred any time between the censoring point and ∞. This is taken into account by taking the expectation over all possible times of bereavement.

Since our dataset only includes twin pairs for which both twins survive past January 1, 1943, this left-truncation has to be taken into account in the likelihood function. We denote the respective truncation age of twin j on January 1, 1943 by t_j^0. With this the survival function is

$$S(t_1,t_2|T_1>t_1^0,T_2>t_2^0,x) = S(t_1,t_2|x)S(t_1^0,t_2^0|x)^{-1}$$

For given functions $\phi,\lambda,\delta_a,\delta_b$, this leads to the following likelihood contribution of a twin pair:

$$L(t_1,t_2,c_1,c_2|x) = [\ \ c_1c_2S(t_1,t_2|x) - c_1(1-c_2)S_{t_2}(t_1,t_2|x) \\ -(1-c_1)c_2S_{t_1}(t_1,t_2|x) + (1-c_1)(1-c_2)S_{t_1,t_2}(t_1,t_2|x)\ \ ] \\ S(t_1^0,t_2^0|x)^{-1}. \tag{2}$$

Here, c_1 and c_2 denote the censoring indicators for T_1 and T_2 and $S_{t_1,t_2}(t_1,t_2|x) = \frac{\partial^2 S(t_1,t_2|x)}{\partial t_1 \partial t_2}$. The functional forms of S, S_{t_1}, S_{t_2} and S_{t_1,t_2} and details of the derivations are presented in Appendix A.4.

To proceed, we now discuss the specifications of the functions $\phi(x), \lambda(t), \delta_a(t-t_k)$ and $\delta_b(t_k,x)$. The baseline hazard (or force of mortality, or duration dependence) function $\lambda(t)$ is specified as $\lambda(t) = e^{\alpha_1 t+\alpha_2 t^2+\alpha_3 t^3}$. This generalizes the Gompertz function $e^{\alpha_1 t}$ which is commonly used in models for high-age mortality and is known to give a reasonably good fit

across a wide range of ages. However, in our setting, it is particularly important to have a flexible functional form, since we aim to estimate the impact of an intermediate event later in life, and the estimate of that impact may be biased if the baseline hazard is misspecified. For this reason we do not impose $\alpha_2 = \alpha_3 = 0$. As a further precautionary measure against misspecification of λ, we allow it to vary across birth cohorts. Early-life health conditions improved over the birth years in our sample, and Gavrilov and Nosov (1985) show that the more recent cohorts faced disproportionally high gains in mortality reductions at higher ages. We therefore allow the parameter vector $(\alpha_1 \ \alpha_2 \ \alpha_3)$ to be different across three different birth cohort intervals: 1873-1899, 1900-1915 and 1916-1930.

Covariate effects enter the hazard through $\phi(x) = e^{\beta' x}$, as is common in proportional-hazard types of models. The function $\delta_a(t - t_k)$ in the bereavement effect function is specified as a piecewise constant function of $t - t_k$. Specifically, $\log \delta_a(t - t_k) = \delta_q^t$, allowing for three time intervals for $t - t_k$, each represented by values $q \in \{1, 2, 3\}$: up to 1 year, 2 to 4 years and after 4 years. (Note that the notation δ_q^t should not be taken to suggest that δ_a depends on t rather than on $t - t_k$.) Furthermore, we take $\log \delta_b(t_k, x) = \delta_l^{age} + \delta^{x\prime} x$, where δ_l^{age} reflects how the bereavement effect depends on the age interval l in which t_k occurs. We allow for three age intervals for t_k, each represented by values $l \in \{1, 2, 3\}$: below 65, 66 to 79 and above 80.

5 Empirical analysis

5.1 Estimation results

The estimation results (with the exception of the vectors of the baseline hazard parameters) are shown in Table 1. Four different model versions are estimated. Models III and IV are comprehensive models whereas Models I and II are restricted models estimated for the purpose of comparing our comprehensive approach to the two approaches used in the two strands of the twin mortality literature discussed in Section 1. In Model I, the only possible dependence between twin lifespans conditional on covariates is generated by the bereavement effect. This model does not allow for systematic unobserved lifetime determinants of the twins. The diametrically opposite approach is represented by Model II which is a correlated frailty model that does not include a bereavement effect. Model

III synthesizes Models I and II. Model IV simply extends Model III by allowing for a more extensive set of determinants of the bereavement effect.

When comparing the estimates of the correlated Gamma frailty distribution in Model II to those from Models III and IV, one finds considerably higher estimates of the variance and the correlation parameters in Model II. This applies to monozygotic $(\sigma^2_{MZ}, \rho_{MZ})$ as well as dizygotic $(\sigma^2_{DZ}, \rho_{DZ})$ twin pairs. The reduction in the estimates of the correlation parameters means that ignoring bereavement effects leads to an overestimation of the importance of genetic and childhood-environmental characteristics as mortality determinants. The estimated correlations in Model II not only reflect the influence of shared genetic and environmental determinants but also capture the causal dependence between twin lifespans due to a bereavement effect.

Perhaps even more importantly, the correlation parameter for monozygotic twins decreases by more than the correlation parameter for dizygotic twins, once we account for a bereavement effect.[9] This means that studies based on correlated frailty models that ignore bereavement effects tend to overestimate the importance of genetic factors in lifetime durations. Below we return to this finding.

The bereavement effect estimates in Model I are enormous. For example, they imply that a monozygotic male twin who is 75 years old and has lost his co-twin at the age of 70 would die on average 2.2 years earlier compared to if he had never experienced this loss. In Models III and IV we find considerably lower estimates (28% lower residual life expectancy in Model IV). This illustrates how strongly the estimates of the bereavement effect are biased if shared unobserved genetic and environmental childhood factors are ignored.

Figure 1 shows the shape of the baseline hazard functions in Model IV for each of the three birth cohort intervals $c = 1, 2, 3$ (1873-1899, 1900-1915, 1916-1930).[10] Evidently, younger cohorts have a considerably lower mortality hazard at higher ages compared to the older cohorts. This change in the aging process over time is known as the late-life mortality deceleration (see Gavrilov and Nosov, 1985). The covariate effects reported in

[9] For this parameter the associated 90% confidence intervals for the estimates in Model II and IV do not overlap.

[10] Since few individuals reached ages above 95, the relevant age period is 55-95. Note that due to left-truncation and right-censoring at ages that vary in the sample, it is not informative to display raw mortality figures.

Table 1 are fully in line with those in the literature, especially in studies using the Danish Twin Registry. We therefore refer to that literature for detailed discussions. As pointed out in this literature, to some extent the estimated zygosity effect in β may be affected by disproportionately high early-life mortality selection among one or both members of MZ twin pairs.

The bereavement effect in Model IV is piecewise constant in the time since bereavement, accounting for three successive time intervals. The three parameter estimates are all significantly positive and of similar size, suggesting persistence of the effect. The size of the estimated bereavement effect for dizygotic twin pairs is only half the size of that of monozygotic twin pairs. This is in line with findings in psychological studies (see Segal and Bouchard, 1993; Segal et al., 1995) based on survey data with bereaved twins. These studies construct measures of grief intensities and find significantly higher grief intensities for monozygotic twins. The finding that bereavement effects are larger for monozygotic twins explains why the estimated frailty correlation parameter for monozygotic twins is so sensitive to whether bereavement effects are taken into account. To put this differently: ignoring that bereavement effects are larger for monozygotic twins leads to an overestimation of the importance of genetic drivers of mortality.

In Model IV, the bereavement effect also depends on the age at bereavement. Evidently, there is a decrease in the bereavement effect in the age at which the loss occurs. In particular, the effect of losing the co-twin at the age of 80 compared to experiencing no bereavement is relatively small, with an implied decrease in residual life expectancy of 0.68 years (Table 2: monozygotic males at age 85). This is also in line with findings from psychological studies with bereaved twins who find a negative correlation between grief intensities and the age at which bereavement is experienced (see Segal et al., 1995).

We performed a range of sensitivity analyses. The fit of the model deteriorates significantly if we impose a simple Gompertz functional form for the baseline hazard $\lambda(t)$. The same applies if we only allow for a second-order polynomial in $\log \lambda(t)$ or if we impose the restriction that the function $\lambda(t)$ does not depend on the birth year. Allowing for a fourth-order polynomial in $\log \lambda(t)$ does not further improve the fit of the model. We tried to estimate the model separately by birth cohort intervals of 10 or 15 years but the resulting subsamples turned out to be too small for reliable inference. Moderate increases of the lowest ages at which lifetimes are left-truncated do not have important implications

for the empirical findings.

5.2 Residual life expectancies

One advantage of our modeling approach at the individual level is that it enables us to predict residual lifetimes as a function of the time at which the loss of the co-twin is experienced. Expected residual lifetimes are relevant for health care policy and are frequently calculated within the demographic and gerontological literature. The expected residual lifetime at age s is computed as (see e.g. Lancaster, 1990)

$$\mu(s,x) = \frac{\int_s^\infty S(t|x)\,dt}{S(s|x)}.$$

Expected residual lifetimes for monozygotic and dizygotic male and female twins, as implied by the estimated Model IV, are presented in Table 2. A male monozygotic twin who has reached the age of 75 and lost his co-twin at the age of 70 will live on average for 6.42 remaining years. If he never experiences twin bereavement, he will live on average for 8.09 years longer. A similar pattern is observed for female twins. Since the dependence of the bereavement effect on sex is insignificant, we set this effect to zero in our calculations for Table 2.

6 Conclusion

The contribution of this paper is twofold. First, we demonstrate that a symmetric duration model with dependent unobserved determinants and a causal effect of one duration on the hazard of the other duration is identified from minimal covariate variation. This model has a wider relevance for the empirical study of parallel systems and networks and for epidemiological research. In the most extreme case, the model allows the two durations to be indistinguishable in terms of observed characteristics. So even if the two durations can not be indexed and the only observable covariates are characteristics of the pair, the identification result applies.

Second, our empirical analysis unites two models that previously have only been used separately in studies analyzing twin lifespans. We thus disentangle the effects of interest in this strand in the literature: the causal effect of bereavement on the one hand,

and the influence of genetic factors and early childhood experiences on the other. This has implications for both strands of literature. We find significantly positive bereavement effects that decrease in magnitude with the age at bereavement and that are more pronounced in monozygotic twin pairs than in dizygotic pairs. If the influence of unobserved time-constant correlated factors is ignored, as in previous studies on twin bereavement, bereavement effects are severely overestimated.

Likewise, ignoring bereavement effects in twin mortality analysis leads to biased results. Our empirical results suggest that studies that aim to shed light on the importance of genetic and early-life environmental characteristics and that ignore bereavement effects tend to overestimate the importance of those characteristics as mortality determinants. Moreover, they tend to overestimate the importance of genetic factors in lifetime durations. The latter is at least partly due to the fact that bereavement effects are larger among monozygotic twins.

We view it as an interesting topic for further research to examine under which conditions models are identified that allow bereavement effects to vary with individual-specific unobserved characteristics. In particular, it may be that such characteristics have genetic determinants, and this would imply that the individual magnitude of the bereavement effect has a genetic component. Even if such genetic determinants do not have direct effects on over-all mortality, they would still have implications for mortality, by way of their indirect pathway through bereavement effects. In this respect, genome-wide association studies may provide interesting insights into which genes are involved.

Table 1: ***Estimation Results:*** *Twin mortality models*

	(Model I)		(Model II)		(Model III)		(Model IV)	
	Bivariate Model with ber.effect		**Corr. Frailty Model no ber.effect**		**Corr. Frailty Model with ber.effect**		**Corr. Frailty Model with ber.eff. (extended)**	
Variable	Estimate	St.Error	Estimate	St.Error	Estimate	St.Error	Estimate	St.Error
Covariates:								
male	.3979***	(.0164)	.5856***	(.0284)	.5143***	(.0291)	.4954***	(.0292)
log(birth year)	-.1088***	(.0239)	-.245***	(.0504)	-.1808***	(.0391)	-.1806***	(.0382)
birth in spring season	.0253	(.0183)	.0447	(.0274)	.0363	(.0237)	.0355	(.0232)
Copenhagen	.0876***	(.024)	.1195***	(.0357)	.1073***	(.0309)	.1059***	(.0304)
dizygotic	.1557***	(.024)	.0435	(.0341)	.1137***	(.035)	.1129***	(.0346)
Bereavement effect:								
first year	.517***	(.0544)	-	-	.4029***	(.0768)	.4303***	(.0884)
second to fourth year	.5211***	(.0398)	-	-	.4037***	(.0699)	.4226***	(.0824)
after four years	.545***	(.0328)	-	-	.4184***	(.0754)	.3944***	(.0878)
dizygotic	-.2798***	(.036)	-	-	-.2538***	(.0737)	-.2491***	(.0776)
male	-	-	-	-	-	-	.0382	(.036)
age at ber. below 65	-	-	-	-	-	-	.1077***	(.0362)
age at ber. above 80	-	-	-	-	-	-	-.1087**	(.0479)
Corr. Gamma frailty:								
variance monozygotic	-	-	.4982***	(.0598)	.3287***	(.0646)	.3171***	(.0658)
dizygotic	-	-	.4697***	(.0581)	.2748***	(.0602)	.2712***	(.0619)
correlation monozygotic	-	-	.945***	(.0638)	.5957***	(.1586)	.5162***	(.1826)
dizygotic	-	-	.4964***	(.0532)	.3452***	(.1356)	.1868	(.1725)

Maximum likelihood parameter estimates based on equation (2). All four estimations are based on 9495 twin pairs (subsample of twin pairs surviving past 1943). Standard errors are reported in parentheses and *, **, or *** indicate a significance level of 0.1, 0.05, or 0.01.

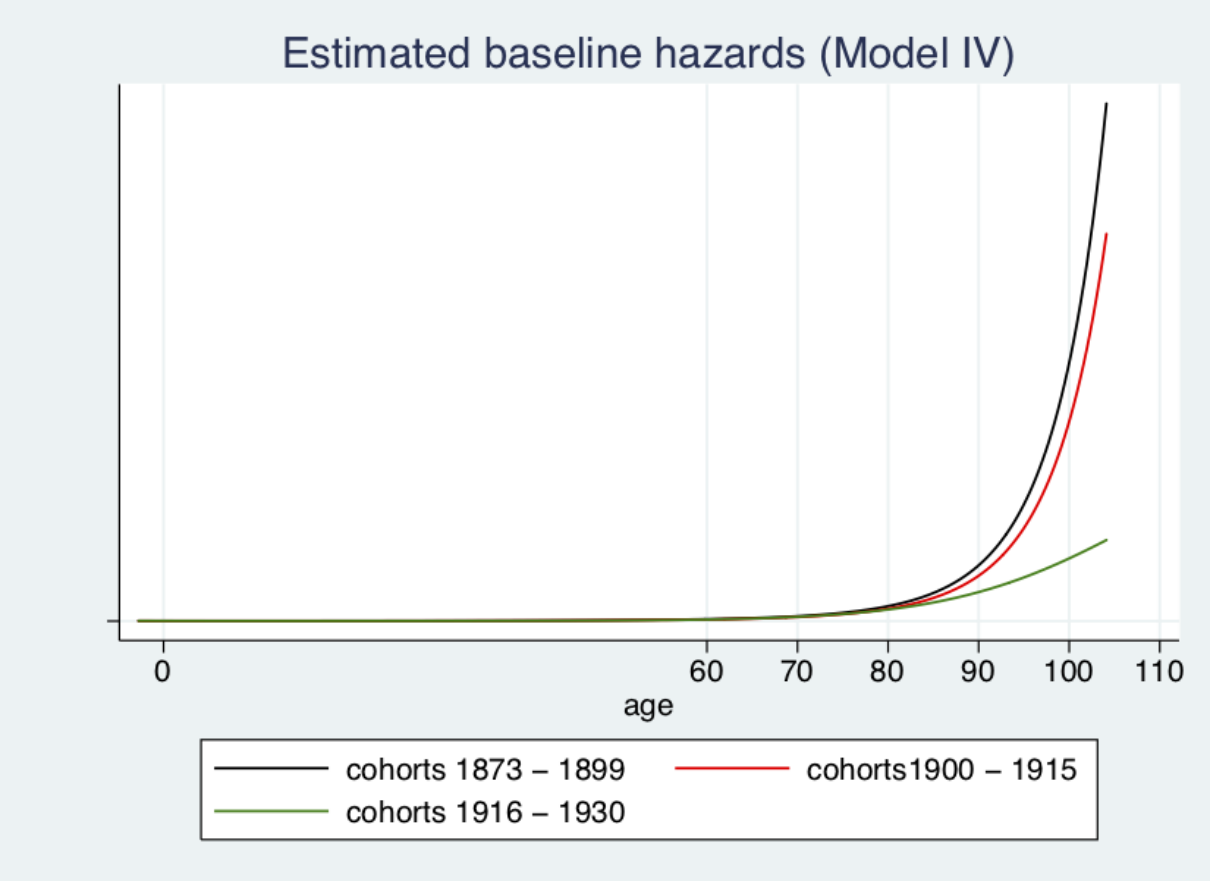

Figure 1: Baseline hazard functions based on estimates of Model IV with a generalization of the Gompertz baseline hazard: $\lambda(t) = \exp(\alpha_1 t + \alpha_2 t^2 + \alpha_3 t^3)$ where the parameter vector $(\alpha_1\ \alpha_2\ \alpha_3)$ is birth-cohort interval specific.

Residual Life Expectancy for Monozygotic Twins

	Male					Female				
Age	No Bereav.	Bereavement at age: 60	70	80	90	No Bereav.	Bereavement at age: 60	70	80	90
65	14.5	11.46				17.73	14.46			
75	8.09	6.00	6.42			10.49	8.06	8.55		
85	3.57	2.45	2.66	2.89		4.99	3.55	3.83	4.13	
95	1.15	.74	.81	.90	.90	1.75	1.14	1.26	1.38	1.38

Residual Life Expectancy for Dizygotic Twins

	Male					Female				
Age	No Bereav.	Bereavement at age: 60	70	80	90	No Bereav.	Bereavement at age: 60	70	80	90
65	13.79	12.26				16.98	15.32			
75	7.59	6.53	6.97			9.92	8.69	9.2		
85	3.29	2.72	2.95	3.20		4.64	3.91	4.21	4.53	
95	1.04	.83	.92	1.01	1.01	1.59	1.29	1.41	1.54	1.54

Table 2: ***Residual life expectancies of twins in years by zygosity and sex.*** Based on estimated Model IV. The first column reports the current age. Subsequent columns report the corresponding residual life expectancies, given that bereavement is never experienced or occurs at ages 60, 70, 80 or 90.

References

Abbring, J. H. and G. J. van den Berg (2003). The nonparametric identification of treatment effects in duration models. *Econometrica 71*(5), 1491–1517.

Ansart, S., C. Pelat, P. Boelle, F. Carrat, A. Flahault, and A. Valleron (2009). Mortality burden of the 1918-1919 influenza pandemic in Europe. *Influenza and Other Respiratory Viruses 3*, 99–106.

Bowling, A. (1987). Mortality after bereavement: A review of the literature on survival periods and factors affecting survival. *Social Science & Medicine 24*(2), 117–124.

Canudas-Romo, V. and A. Erlangsen (2008). Denmark: the lowest excess mortality during the influenza pandemic of 1918, Working paper, University of Aarhus.

Elbers, C. and G. Ridder (1982). True and spurious duration dependence: The identifiability of the proportional hazard model. *Review of Economic Studies 49*(3), 403–409.

Flournoy, N. (1992). Discussion of assessment of dependence in the life times of twins. In J. Klein and P. Goel (Eds.), *Survival Analysis: State of the Art*, pp. 91–97. Kluwer Academic Publishers.

Gavrilov, L. A. and V. N. Nosov (1985). A new trend in human mortality decline: derectangularization of the survival curve. *Age 8*, 93.

Gourieroux, C. and Y. Lu (2015). Love and death: A Freund model with frailty. *Insurance: Mathematics and Economics 63*, 191–203.

Hauge, M., B. Harvald, M. Fischer, K. Jensen, J. Gotlieb, J. Reabild, R. Shapiro, and T. Videbech (1968). The Danish twin register. *Acta Geneticae Medicae et Gemillologiae 17*(1968), 315–334.

Herskind, A. M., M. McGue, N. V. Holm, T. I. A. Sørensen, B. Harvald, and J. W. Vaupel (1996). The heritability of human longevity: A population-based study of 2872 Danish twin pairs born 1870–1900. *Human Genetics 97*(3), 319–323.

Holm, N. V. (1983). *The use of twin studies to investigate causes of diseases with complex etiology with a focus on cancer*. Ph. D. thesis, Odense University.

Hougaard, P., B. Harvald, and N. V. Holm (1992a). Assessment of dependence in the life times of twins. In J. Klein and P. Goel (Eds.), *Survival Analysis: State of the Art*, pp. 77–90. Kluwer Academic Publishers.

Hougaard, P., B. Harvald, and N. V. Holm (1992b, March). Measuring the similarities between the lifetimes of adult Danish twins born between 1881-1930. *Journal of the American Statistical Association 87*(417), 17–24.

Kortram, R. A., A. C. M. van Rooij, A. J. Lenstra, and G. Ridder (1995). Constructive identification of the mixed proportional hazards model. *Statistica Neerlandica 49*(3), 269–281.

Lancaster, T. (1990). *The Econometric Analysis of Transition Data*. Cambridge: Cambridge University Press.

Lichtenstein, P., M. Gatz, and S. Berg (1998, 5). A twin study of mortality after spousal bereavement. *Psychological Medicine 28*, 635–643.

Lindeboom, M., F. Portrait, and G. J. van den Berg (2002). An econometric analysis of the mental-health effects of major events in the life of older individuals. *Health Economics 11*(6), 505–520.

Lu, Y. (2017). Broken-heart, common life, heterogeneity: Analyzing the spousal mortality dependence. *ASTIN Bulletin 47*(3), 837–874.

Lykken, D. T. (1978). The diagnosis of zygosity in twins. *Behavior Genetics 8*, 437–473.

Manor, O. and Z. Eisenbach (2003). Mortality after spousal loss: are there socio-demographic differences? *Social Science & Medicine 56*(2), 405 – 413.

Sanders, C. M. (1980). A comparison of adult bereavement in the death of a spouse, child, and parent. *Journal of Death and Dying 10*(4), 303–322.

Sanders, C. M. (1999). *Grief: The Mourning After: Dealing with Adult Bereavement, 2nd Edition*. Wiley & Sons, Chichester.

Segal, N. L. and S. A. Blozis (2002). Psychobiological and evolutionary perspectives on coping and health characteristics following loss: a twin study. *Twin Research 5*, 175–87.

Segal, N. L. and T. J. Bouchard (1993). Grief intensity following the loss of a twin and other relatives: test of kinship genetic hypotheses. *Human Biology 65*, 87–105.

Segal, N. L. and S. L. Ream (1998). Decrease in grief intensity for deceased twin and non-twin relatives: an evolutionary perspective. *Personality and Individual Differences 25*(2), 317 – 325.

Segal, N. L., L. J. Sussman, W. D. Marelich, J. Mearns, and S. A. Blozis (2002, 6). Monozygotic and dizygotic twins' retrospective and current bereavement-related behaviors: An evolutionary perspective. *Twin Research 5*, 188–195.

Segal, N. L., S. M. Wilson, T. J. Bouchard, and D. G. Gitlin (1995). Comparative grief experiences of bereaved twins and other bereaved relatives. *Personality and Individual Differences 18*(4), 511 – 524.

Selye, H. (1936). A syndrome produced by diverse nocuous agents. *Nature 138*, 32.

Selye, H. (1955). Stress and disease. *Science 122*(3171), 625–631.

Skytthe, A., K. Kyvik, N. Holm, J. Vaupel, and K. Christensen (2002). The Danish twin registry: 127 birth cohorts of twins. *Twin Research 5*, 352–357.

Stroebe, M. S. and W. Stroebe (1987). *Bereavement and Health: The Psychological and Physical Consequences of Partner Loss*. Cambridge University Press.

Stroebe, M. S., W. Stroebe, and R. O. Hansson (1993). The mortality of bereavement: A review. In *Handbook of Bereavement: Theory, Research and Intervention*, pp. 175 – 195. Cambridge University.

Tomassini, C., F. C. Billari, A. Rosina, A. Skytthe, and K. Christensen (2001). Born together-die together. live together-die together. the role of the partner and of the co-twin on longevity at very old ages. *Genus 57*(3-4), 63–82.

Tomassini, C., A. Rosina, F. C. Billari, A. Skytthe, and K. Christensen (2002). The effect of losing the twin and losing the partner on mortality. *Twin Research 5*, 210–217.

Van den Berg, G. J. (2001). Duration models: specification, identification and multiple durations. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 5. Elsevier.

Van den Berg, G. J., M. Lindeboom, and F. Portrait (2011). Conjugal bereavement effects on health and mortality at advanced ages. *Journal of Health Economics 30*(4), 774–794.

Widder, D. V. (1946). *The Laplace Transform*. Princeton University Press, Princeton.

Wienke, A., K. Christensen, A. Skytthe, and A. I. Yashin (2002). Genetic analysis of cause of death in a mixture model of bivariate lifetime data. *Statistical Modelling 2*, 89–102.

Wienke, A., N. Holm, A. Skytthe, and A. Yashin (2001, August). The heritability of mortality due to heart diseases: a correlated frailty model applied to Danish twins. *Twin Research 4*, 266–74.

Woodward, J. (1988). The bereaved twin. *Acta Geneticae Medicae et Gemellologiae 37*, 173–80.

Yashin, A. I. and I. A. Iachine (1995a). Genetic analysis of durations: Correlated frailty model applied to survival of Danish twins. *Genetic Epidemiology 12*, 529–538.

Yashin, A. I. and I. A. Iachine (1995b). How long can humans live? lower bound for biological limit of human longevity calculated from Danish twin data using correlated frailty model. *Mechanisms of Ageing and Development 80*, 147–169.

Appendix

A.1 Identification of λ and ϕ

The survival function of $Z|x$ with $Z = \min\{T_1, T_2\}$ is derived as follows

$$\begin{aligned} P(Z > t|x) &= P(T_1 > t, T_2 > t|x) \\ &= \int_0^\infty \int_0^\infty P(T_1 > t|x, V_1) P(T_2 > t|x, V_2), dG(v_1, v_2) \\ &= \int_0^\infty \int_0^\infty e^{-\phi(x)\Lambda(t)(V_1+V_2)}, dG(v_1, v_2) \\ &= \int_0^\infty e^{-\phi(x)\Lambda(t)W}, dG_W(w) \qquad \text{with } W = V_1 + V_2. \end{aligned} \tag{3}$$

For the second equality we exploit that before the first death has occurred no bereavement effect is experienced yet. Consequently, conditional on x and V the events $(T_1 > t)$ and $(T_2 > t)$ are independent. We further use Assumption 4 which implies $G(v_1, v_2|x) = G(v_1, v_2)$.

The distribution of Z has a hazard rate of the mixed proportional form: $\theta_z(t|x, W) = \theta(t|T_2 \geq t, x, V_1) + \theta(t|T_1 \geq t, x, V_2) = \lambda(t)\phi(x)W$ with frailty $W = V_1 + V_2$ drawn from distribution G_W. The results by Elbers and Ridder (1982) (see also Lancaster, 1990; Van den Berg, 2001, for an overview) on the identification of the mixed proportional hazard model imply that, under Assumptions 1-4, the model in Equation (3), characterized by the functions λ, ϕ and G_W, is identified. In particular, Assumption 1 assures sufficient covariate variation in form of at least one dummy variable.[11] Further, we require the distribution of W to be independent of x and to have a positive and finite mean. Assumption 4 assures the independence of (V_1, V_2) and x. From this the independence of $W = V_1 + V_2$ directly follows. Similarly, as V_1 and V_2 are assumed to have finite positive mean, so does W.

[11] See also Kortram et al. (1995) for the case of only two possible values for $\phi(x)$.

A.2 Identification of δ_a and δ_b

We consider the following hazard rate,

$$\theta(t|T_k = 0, x, V_j) = \tilde{\lambda}_j(t)\tilde{\phi}_j(x)V_j \text{ with } \tilde{\lambda}_j(t) = \lambda(t)\delta_a,\ \tilde{\phi}_j(x) = \phi(x)\delta_b(0, x), \tag{4}$$

where the frailties V_j are drawn from $G_{V_j|T_k=0,x}$ for $j, k \in \{1, 2\}$ and $j \neq k$. This hazard rate can be said to have a mixed proportional form as it is proportional in t, x, and an unobserved frailty term.

We now demonstrate that the conditional frailty distribution $G_{V_j|T_k=0,x}$ does not depend on x. Its density is given by:

$$\begin{aligned} f(v_j|T_k = 0, x) &= \frac{\theta_k(0|x, V_j)S_k(0|x, V_j)f(v_j|x)}{\theta_k(0|x)S_k(0|x)} \\ &= \frac{\int_0^\infty \lambda(0)\phi(x)v_k \,\mathrm{d}G(v_k|x, V_j)f(v_j|x)}{\int_0^\infty \lambda(0)\phi(x)v_k \,\mathrm{d}G(v_k|x)} \\ &= \frac{E(V_k|x, V_j)f(v_j|x)}{E(V_k|x)}. \end{aligned} \tag{5}$$

According to Assumption 4 (V_1, V_2) are independent of x. Therefore, Equation (5) simplifies to

$$f(v_j|T_k = 0, x) = \frac{E(V_k|V_j)f(v_j)}{E(V_k)}. \tag{6}$$

Note that the right hand side of (6) does not depend on x. From Equation (6) it also follows that the distribution of $(V_j|T_k = 0)$ for $j, k \in \{1, 2\}$ and $j \neq k$ has a positive and finite mean, since $G(v_1, v_2)$ has this property.

Assumption 2 states that the functions $\phi(x)$ and $\delta_b(0, x)$ are not proportional, assuring that the function $\hat{\phi}(x) = \phi(x)\delta_b(0, x)$ generates sufficient exogenous variation.

Thus, following Elbers and Ridder (1982), under Assumptions 1-4, the mixed proportional hazard model defined by $\{\tilde{\lambda}, \tilde{\phi}, G_{V_j|T_k=0,x}\}$ is identified. Since λ is known from Appendix A.1, this in turn identifies δ_a. Similarly, since ϕ is known from Appendix A.1, the function $\delta_b(0, x)$ follows as a function of x.

To identify $\delta_b(t, x)$ as a function of t and x, we exploit information on the jump of the

hazard rate at the moment of bereavement

$$\frac{\lim_{s\downarrow t}\theta(s|T_k=t,x)}{\theta(t|T_k=t,x)} = \frac{\phi(x)\delta_b(t,x)\lim_{s\downarrow t}\delta_a(s-t)\lambda(s)E(V_j|T_j\geq s,T_k=t,x)}{\phi(x)\lambda(t)E(V_j|T_j\geq t,T_k=t,x)}$$
$$= \delta_b(t,x)\lim_{s\downarrow t}\delta_a(s-t)\frac{\lim_{s\downarrow t}\lambda(s)}{\lambda(t)}. \tag{7}$$

Assumptions 2 and 3 assure the existence of $\lim_{s\downarrow t}\delta_a(s-t)$ and $\lim_{s\downarrow t}\lambda(s)$. Accordingly, the second equality directly follows from Assumption 4, stating that $\lim_{s\downarrow t}E(V_j|T_j \geq s,T_k=t,x) = E(V_j|T_j\geq t,T_k=t,x)$. Note, that the left hand side of Equation 7 is observable for all $(t,x)\in(0,\infty)\times\mathcal{X}$. Since $\lim_{s\downarrow t}\delta_a(s-t)$, $\lim_{s\downarrow t}\lambda(s)$ and $\lambda(t)$ are known from previous steps, we can trace out the function $\delta_b(t,x)$ over $(0,\infty)\times\mathcal{X}$.

Together, Appendix A.1 and Appendix A.2 thus prove Proposition 1.

A.3 Proof of Proposition 2

. Recall that the functions $\lambda,\phi,\delta_a,\delta_b$ in Model 1 are identified under Assumptions 1-4. The only function that remains undetermined is the bivariate frailty distribution G. For this we adopt the additional Assumption 5.

The observable density $f(t_1,t_2|x)$ for $t_1>t_2$ can be expressed as follows

$$\begin{aligned} f(t_1,t_2|x) &= \int_0^\infty\int_0^\infty f(t_1|T_2=t_2,x,V_1)f(t_2|x,V_2)\,\mathrm{d}G(v_1,v_2) \\ &= c(t_1,t_2,x)\int_0^\infty\int_0^\infty V_1V_2e^{-\phi(x)(\Lambda(t_2)+\Delta_1(t_1,t_2,x))V_1}e^{-\phi(x)\Lambda(t_2)V_2}\,\mathrm{d}G(v_1,v_2) \\ &= c(t_1,t_2,x)\partial^2_{s_1,s_2}\mathcal{L}_G\big(\phi(x)(\Lambda(t_2)+\Delta_1(t_1,t_2,x)),\phi(x)\Lambda(t_2)\big), \end{aligned}$$

with $c(t_1,t_2,x) = \lambda(t_1)\lambda(t_2)\phi(x)^2\delta_a(t_1-t_2)\delta_b(t_2,x)$, $\Delta(t_1,t_2,x) = \tilde{\Lambda}(t_1,t_2)\delta_b(t_2,x)$ and bivariate Laplace transform $\mathcal{L}_G$ with cross derivative $\partial^2_{s_1,s_2}\mathcal{L}_G$.

Absolute monotonicity and complete monotonicity:

Definition 1. *Let Ω be a nonempty open set in $\mathbb{R}^n$. A function $f:\Omega\to\mathbb{R}$ is absolutely monotone if it is nonnegative and has nonnegative continuous partial derivatives of all orders. f is completely monotone if $f\circ m$ is absolutely monotone, where*

$m : x \in \{\omega \in \mathbb{R}^n : -\omega \in \Omega\} \to -x$.[12]

This definition states that a function f is completely monotone if it's derivatives of all orders exist, and if these derivatives are continuous and have switching signs for each order (starting with a positive first derivative). It follows directly that if a function f is completely monotone then all derivatives of second order of f will also be completely monotone. Since the bivariate Laplace transform $\mathcal{L}_G(s_1, s_2)$ is known to be a completely monotone function, it directly follows from Definition 1 that the cross derivative of $\mathcal{L}$ given by $\partial^2_{s_1,s_2}\mathcal{L}_G(s_1, s_2) = \frac{\partial^2 \mathcal{L}_G(s_1,s_2)}{\partial s_1 \partial s_2}$ is also completely monotone.

Tracing out the Laplace transform: The function $f : \mathbb{R}^2_+ \to \mathbb{R}^2_+$ is given by $f(t_1, t_2) = (\phi(x)(\Lambda(t_2)+\Delta(t_1, t_2, x)), \phi(x)\Lambda(t_2))$. It maps the vector (t_1, t_2) on the vector of arguments of the Laplace transform (s_1, s_2), with $s_1 = \phi(x)(\Lambda(t_2) + \Delta(t_1, t_2, x))$ and $s_2 = \phi(x)\Lambda(t_2)$. In the following we will show that we can vary (t_1, t_2) on an open set such that $f(t_1, t_2)$ will also attain all values in a nonempty open set. Under Assumption 5 (with $t_1 > t_2 \ \forall$ $(t_1, t_2) \in \Psi$) it holds that at all points (t_1, t_2) in the open set Ψ the first derivatives of f exist and are continuous and f has Jacobian

$$J_f(t_1, t_2) = \begin{bmatrix} \phi(x)\lambda(t_1)\delta(t_1, t_2, x) & \phi(x)(\lambda(t_2) + \frac{\partial\Delta(t_1,t_2,x)}{t_2}) \\ 0 & \phi(x)\lambda(t_2) \end{bmatrix}.$$

Note, that the determinant of J_f is given by $det(J_f(t_1, t_2)) = \phi(x)^2\lambda(t_1)\lambda(t_2)\delta_1(t_1, t_2, x)$, and since under Assumptions 1-4 the functions $\phi, \lambda, \delta_a, \delta_b$ can only attain strictly positive (and finite) values on Ψ, it follows that $det(J_f(t_1, t_2)) \neq 0 \ \forall \ (t_1, t_2) \in \Psi$. Assumption 5 assures that $\frac{\partial\Delta(t_1,t_2,x)}{t_2}$ exists and is continuous on Ψ. Therefore, on the nonempty open set Ψ the function $f(t_1, t_2)$ is continuously differentiable with invertible Jacobian J_f. Using the inverse-function theorem it follows that there exists an nonempty open set $\Upsilon \subset (0, \infty)^2$ such that the function $f(t_1, t_2)$ attains all values in Υ when t_1 and t_2 vary over $\Psi \subset (0, \infty)^2$.

□

[12]For $n = 1$ this definition reduces to the familiar definitions in Widder (1946).

A.4 Derivation of the likelihood function

In the following the functional forms of S, S_{t_1}, S_{t_2} and S_{t_1,t_2} are derived. We start with the survival function $S(t_1, t_2|x) = P(T_1 > t_1, T_2 > t_2|x)$:

$$S(t_1, t_2|x) = \begin{cases} S^*(t_1, t_1|x) - \int_{t_2}^{t_1} S_{t_2}(t_1, \tau|x)\, d\tau & \text{, for } t_1 \geq t_2 \\ S^*(t_2, t_2|x) - \int_{t_1}^{t_2} S_{t_1}(\tau, t_2|x)\, d\tau & \text{, for } t_1 < t_2 \end{cases}$$

Here, $S^*(t_1, t_2|x)$ denotes the survival function in the absence of a bereavement effect

$$\begin{aligned} S^*(t_1, t_2|x) &= \iint_0^\infty P(T_1 > t_1|x, V_1) P(T_2 > t_2|x, V_2)\, \mathrm{d}G(v_1, v_2) \\ &= \iiint_0^\infty \mathrm{e}^{\phi(x)\Lambda(t_1)(\tilde{V}_0+\tilde{V}_1)} \mathrm{e}^{\phi(x)\Lambda(t_2)(\tilde{V}_0+\tilde{V}_2)}\, \mathrm{d}G(\tilde{v}_0) dG(\tilde{v}_1) dG(\tilde{v}_2) \\ &= \int_0^\infty \mathrm{e}^{\phi(x)[\Lambda(t_1)+\Lambda(t_2)]\tilde{V}_0}\, \mathrm{d}G(\tilde{v}_0) \int_0^\infty \mathrm{e}^{\phi(x)\Lambda(t_1)\tilde{V}_1}\, \mathrm{d}G(\tilde{v}_1) \int_0^\infty \mathrm{e}^{\phi(x)\Lambda(t_2)\tilde{V}_2}\, \mathrm{d}G(\tilde{v}_2) \\ &= (1 + \sigma^2\phi(x)[\Lambda(t_1) + \Lambda(t_2)])^{-\frac{\rho}{\sigma^2}} (1 + \sigma^2\phi(x)\Lambda(t_1))^{-\frac{(1-\rho)}{\sigma^2}} (1 + \sigma^2\phi(x)\Lambda(t_2))^{-\frac{(1-\rho)}{\sigma^2}}. \end{aligned}$$

The last three equalities follow from the assumption that $G(v_1, v_2)$ is a Cherian bivariate Gamma distribution, i.e. the terms $\tilde{V}_0, \tilde{V}_1, \tilde{V}_2$ are independent and drawn from univariate Gamma distributions: $\tilde{V}_0 \sim \Gamma(\rho\sigma^{-2}, \sigma^{-2})$ and $\tilde{V}_1, \tilde{V}_2 \sim \Gamma((1-\rho)\sigma^{-2}, \sigma^{-2})$.

In the following S_{t_j} is derived. For this purpose we define the functions g_a, g_b and g_c

$$\begin{aligned} g_a(s_1, s_2, x) &= 1 + \sigma^2\phi(x)[\Lambda(s_2) + \Delta(s_1|s_2, x)] \\ g_b(s_1, s_2, x) &= 1 + \sigma^2\phi(x)[2\Lambda(s_2) + \Delta(s_1|s_2, x)] \\ g_c(s, x) &= 1 + \sigma^2\phi(x)\Lambda(s). \end{aligned}$$

with $\Delta(s_1|s_2, x) = \int_{s_2}^{s_1} \lambda(u)\delta_a(u - s_2)\delta_b(s_2, x)\ \mathrm{d}u$.

We can now derive $S_{t_j}(t_j, t_k|x) = \frac{\partial S(t_j, t_k|x)}{\partial t_j} = -P(T_j = t_j, T_k > t_k|x)$. Let $t_j \geq t_k$ with

$j,k \in \{1,2\}, j \neq k$

$$
\begin{aligned}
S_{t_k}(t_j,t_k|x) &= \iint_0^\infty P(T_j > t_j|T_k = t_k, x, V_j) P(T_k = t_k|x, V_k)\, \mathrm{d}G(v_j, v_k) \\
&= \phi(x)\lambda(t_k) \\
&\quad \iiint_0^\infty (\tilde{V}_0 + \tilde{V}_k) \mathrm{e}^{\phi(x)[\Lambda(t_k)+\Delta(t_j|t_k,x)](\tilde{V}_0+\tilde{V}_j)} \mathrm{e}^{\phi(x)\Lambda(t_k)(\tilde{V}_0+\tilde{V}_k)}\, \mathrm{d}G(\tilde{v}_0) dG(\tilde{v}_j) dG(\tilde{v}_k) \\
&= \phi(x)\lambda(t_k) g_b(t_j,t_k,x)^{-(\frac{\rho}{\sigma^2}+1)} g_c(t_k,x)^{-(\frac{(1-\rho)}{\sigma^2})} g_a(t_j,t_k,x)^{-(\frac{(1-\rho)}{\sigma^2}+1)} \\
&\quad [\rho g_a(t_j,t_k,x) + (1-\rho) g_b(t_j,t_k,x)].
\end{aligned}
$$

This yields

$$
S_{t_j}(t_j,t_k|x) = \begin{cases}
\frac{\partial S^*(t_j,t_j|x)}{\partial t_j} + \int_{t_k}^{t_j} S_{t_1,t_2}(t_1,\tau|x)\, d\tau & , \text{for } t_j > t_k \\
\phi(x)\lambda(t_k) g_b(t_j,t_k,x)^{-(\frac{\rho}{\sigma^2}+1)} g_c(t_k,x)^{-(\frac{(1-\rho)}{\sigma^2})} & \\
g_a(t_j,t_k,x)^{-(\frac{(1-\rho)}{\sigma^2}+1)} [\rho g_a(t_j,t_k,x) + (1-\rho) g_b(t_j,t_k,x)] & , \text{for } t_j \leq t_k.
\end{cases}
$$

Finally, $S_{t_1,t_2}(t_1,t_2|x) = \frac{\partial^2 S(t_1,t_2|x)}{\partial t_1 \partial t_2} = P(T_1 = t_1, T_2 = t_2|x) = f^*(\max\{t_1,t_2\}, \min\{t_1,t_2\})$ with

$$
\begin{aligned}
f^*(t_j,t_k) = \phi(x)^2 \lambda(t_j)\lambda(t_k)\delta_a(t_j - t_k)\delta_b(t_k,x) & \\
g_b(t_j,t_k,x)^{-(\frac{\rho}{\sigma^2}+2)} g_a(t_j,t_k,x)^{-(\frac{(1-\rho)}{\sigma^2}+1)} g_c(t_k,x)^{-(\frac{(1-\rho)}{\sigma^2}+1)} & \\
[\rho(\rho+\sigma^2) g_a(t_j,t_k,x) g_c(t_k,x) + \rho(1-\rho) g_b(t_j,t_k,x) g_c(t_k,x) & \\
\rho(1-\rho) g_b(t_j,t_k,x) g_a(t_j,t_k,x) + (1-\rho)^2 g_b(t_j,t_k,x)^2]. &
\end{aligned}
$$

In the estimation, the integrals $\int_{t_2}^{t_1} S_{t_2}(t_1,\tau|x)\, d\tau$ and $\int_{t_1}^{t_2} S_{t_1}(\tau,t_2|x)\, d\tau$ are evaluated using numerical integration methods.