Racial disparities in the life lived without parents: A multistate life table approach

Michael Jongho Moon[*] Benjamin-Samuel Schlüter[*] Monica Alexander[*†]

1 Introduction

In the United States, disparities in mortality risk and life expectancy have long existed. In particular, the non-Hispanic black population experienced excessive premature deaths compared to the non-Hispanic white population throughout the twentieth century (Hummer and Chinn, 2011; Dwyer-Lindgren et al., 2022; Caraballo et al., 2023). The historically continuing disparities in life expectancy and premature mortality suggests that their children are likely to face a parental loss at an earlier age. Many have studied and reported life-long health and social repercussions from the traumatic experience in the literature (Stroebe et al., 2007; Niederkrotenthaler et al., 2012; Patterson et al., 2020; Li et al., 2014; Enrico Debiasi and Eriksson, 2021). These disadvantages would exacerbate the existing disparities.

While the racial disparity in life expectancy is commonly investigated in existing literature (Hummer and Chinn, 2011; Dwyer-Lindgren et al., 2022), few studies focus on the resulting differences in parent mortality among the offspring. Umberson et al. (2017) studied the racial differences in exposure to the death of family members at different ages using the Health and Retirement Study and the National Longitudinal Survey of Youth 1997. They compare the age-specific risk of specific family member deaths and cumulative exposures to deaths of multiple members using parametric models.

Our work further contributes to this area focusing specifically on parental bereavements. We use multiple sources of public data including population life tables, historical fertility and mortality rates, and survey data with age-specific parent mortality status. The survey data allow statistical estimation of the risk of losing each parent in isolation as well as in order. We also use classic demographic models to supplement the parental loss risk estimation. We combine the estimates in a Bayesian model to measure risks of parental loss and life years lived without parents. We applied the framework to different racial and ethnic groups – namely, Hispanics, non-Hispanic blacks, and non-Hispanic whites – in the United States, 2020.

In the following Method section, we present our framework for constructing the multistate life tables. Results section provides a summary of results comparing the racial and ethnic groups. After presenting the results, we discuss the implications of the findings and planned future work in Discussion section.

[*]Department of Statistical Sciences, University of Toronto
[†]Department of Sociology, University of Toronto

2 Method

In this section, we describe our framework to construct the multistate life tables by parent mortality state including the methods and the sources of data used for the 2020 United States population. We first define the multistate model and outline the multistate life table formulation from the single-state life tables of the 2020 United States population (Arias and Xu, 2022) in Section 2.1. The formulation requires age-specific parent mortality rates and proportions for each state. We use a hierarchical Bayesian model to estimate these quantities using survey data estimates and projections from demographic matrix models. We provide an overview of the data used for the model in Section 2.2 and define the Bayesian model in Section 2.3.

2.1 The multistate life table by parent mortality status

We propose a framework for constructing multistate life tables by parent mortality status from a single-state life table. We first define the multistate model of parental loss as shown in Figure 1. The model considers only the biological parents and assumes that both parents are alive at birth. There are four possible states before death of the focal person: State (1) representing having both parents alive, State (2) having lost their mother only, State (3) having lost their father only, and State (4) having lost both. The model captures all possible transitions between the states before death.

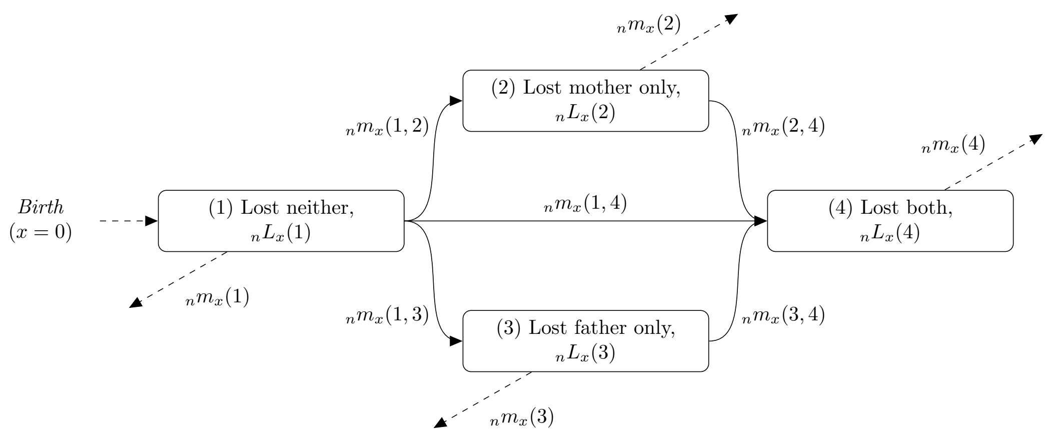

Figure 1. The multistate model of parental loss. ${}_nL_x(i)$ denotes the number of person-years lived in State i, ${}_nm_x(i,j)$ the rate of transitions from State i to j, and ${}_nm_x(i)$ the mortality rate from in State i among life table persons whose age is in $[x, x+n)$.

In a single-state life table, ${}_nL_x$ denotes the number of person-years lived among those whose ages are in $[x, x+n)$. In the multistate life table, we denote the corresponding value of the number of person-years lived in State i among those whose ages are in $[x, x+n)$ with ${}_nL_x(i)$ for $i = 1, 2, 3, 4$. Because the states are mutually exclusive and collectively exhaustive, we have ${}_nL_x = \sum_{i=1}^{4} {}_nL_x(i)$ for any pairs of x and n. For each State i, we denote the proportion of person-years lived in the state of the population aged between x and $x+n$ with ${}_np_x^L(i)$. The superscript L is used to distinguish the proportions from the set of proportions for mortality counts we later introduce. We assume we observe the proportions in the population allowing us to compute the multistate person-years lived based on single-state life tables as shown in Equation 1.

$$ {}_nL_x(i) = {}_np_x^L(i) \cdot {}_nL_x \quad \text{for } i = 1, 2, 3, 4 \tag{1} $$

Similar to the assumption on the multistate person-years, we assume that the transition rates in the life table are consistent with the rates observed in the population. ${}_{n}m_{x}(i,j)$ denotes the transition rate from State i to State j among those whose age is in $[x, x+n)$. We use the observed transition rates and the multistate person-years lived numbers to compute the number of transitions between states in the multistate life table. ${}_{n}d_{x}(i,j)$ denotes the number of transitions from State i to State j among those whose ages are in $[x, x+n)$ and compute the values using Equation 2.

$$ {}_{n}d_{x}(i,j) = {}_{n}m_{x}(i,j) \cdot {}_{n}L_{x} \quad \text{for } i = 1,2,3,4 \tag{2} $$

For example, ${}_{5}d_{10}(1,2)$ is the number of life table persons who lose their mothers by age 15 among those whose both parents were alive at age 10. ${}_{5}m_{10}(1,2)$ is the corresponding rate per person-years lived with both parents between ages 10 and 15. For all infeasible transitions, such as moving from State 2 to State 3, ${}_{n}m_{x}(i,j) = {}_{n}d_{x}(i,j) = 0$.

We compute the mortality numbers based on observed population proportions in the same manner as we compute the multistate person-years. We denote the mortality counts from State i among those whose ages are between $[x, x+n)$ with ${}_{n}d_{x}(i)$. The corresponding total mortality count from all states found in single-state life tables is denoted ${}_{n}d_{x}$. We denote the proportion of deaths from State i out of ${}_{n}d_{x}$ using ${}_{n}p_{x}^{D}(i)$. Assuming the observed population proportions are consistent with the life table proportions, we can compute the multistate mortality counts from the population mortality counts as shown in Equation 3.

$$ {}_{n}d_{x}(i) = {}_{n}p_{x}^{D}(i) \cdot {}_{n}d_{x} \quad \text{for } i = 1,2,3,4 \tag{3} $$

We note that this assumes that those focal persons in different parent mortality states experience the same mortality rates.

We then follow the relationships defined in Schoen (1988) to compute the remaining quantities for the multistate life tables. Namely, we compute the expected life remaining in State i above age x, $e_{x}(i)$ and the number of life table survivors in State i at age x, $\ell_{x}(i)$.

Arias and Xu (2022) published complete period life tables for United States, 2020. Downloadable files were available for each sub-population by Hispanic origin and race. Each table included single-year age groups up to 100 and those above 100 in a single group. We extracted life tables for Hispanic, non-Hispanic black, and non-Hispanic white populations. Specifically, we extracted and aggregated person-year lived, ${}_{5}L_{x}^{r}$, life-table survivors, ${}_{5}\ell_{x}^{r}$, and life-table person deaths, ${}_{5}d_{x}^{r}$, for ages $x \in \{0, 5, ..., 60\}$, and ${}_{\infty}L_{65}$, ${}_{\infty}\ell_{65}$, and ${}_{\infty}d_{65}$ to close the life tables.$r \in \{1,2,3\}$ represents the race/ethnic groups. In the following section, we describe our approach to estimate the parent mortality status proportions and parent mortality rates to propagate the single-state life tables to the multistate life tables.

2.2 Data

We used 2021 Survey of Income and Program Participation (SIPP) public-use data files (United States Census Bureau, 2023) to extract age-specific parent mortality estimates. SIPP is a nationally representative survey that provides social and economic information about the United States population including the family context of individuals. In particular, the survey asked whether the participant's parents were alive or deceased as well as the participant's age at the time of losing each parent if either of them were deceased. The responses provided information about parent mortality status from each respondent's birth to the survey reference year, 2020. While a period life table is constructed based on the mortality statistics of a reference year, we used up to 20 years of each respondent's life information to estimate age-specific parent mortality statistics. For example, a respondent who was 19 years old in 2020 would contribute to mortality statistics for ages 0 to 19 while a respondent who was 49 would contribute to those for ages 30 to 49. The use of retrospective information allowed inclusion of more samples and avoided an excessive number missing estimates. Using the reference year information only would lead to missing a large number of rate estimates, especially for childhood ages. We recognize that the approach may introduce bias with respect to the true period parent mortality statistics. We placed the 20-year limit to contain any bias introduced. We assume parent mortality statistics has not changed significantly between 2021 and 2020 at the population level. We computed standard errors for each statistic following the 2021 SIPP Users' Guide (United States Census Bereau, 2023).

To supplement the survey data, we followed Caswell et al. (2023) to build kinship matrix models to project age-specific parent mortality estimates for each parent independently. The model takes historical fertility and mortality rates to project parent survival and mortality in the projection year. We used historical fertility and mortality data from years 1990 to 2020 collected by Schlüter et al. (2024). The historical data also do not have sampling distributions associated with them. However, we assume that they are observations of latent stochastic processes and estimate the standard errors following the bootstrapping method implemented in Schlüter et al. (2024).

2.3 Bayesian model of parent mortality status proportions and rates

Propagating the multistate life table from a single-state life table requires the age-specific proportions in each parent mortality state and parent mortality rates. We assume that those in different parent mortality states experience the same mortality likelihoods. This allows us to use one proportion estimate to propagate both the person-year lived and the mortality counts.

We estimate the parent mortality proportions and rates by combining direct estimates from the survey data and parent mortality matrix projections in a hierarchical Bayesian model. While the survey data provide direct estimates, they may contain missing data, bias, and/or high variability. The matrix model projects for each parent provide parent mortality estimates that are independent of the other parent. We denote the direct estimates from the survey data with ${}_5p^r_x(i)$ and ${}_5m^r_x(i,j)$ for $i,j \in \{1,2,3,4\}$, $x \in \{0,5,...,60\}$, and $r \in \{1,2,3\}$ respectively for the proportions and rates. We use ${}_5\hat{\mathscr{p}}^r_x(2)$ and ${}_5\hat{\mathscr{p}}^r_x(3)$ to denote the estimated proportion of person-years lived without their mother and without their father respectively; and ${}_5\hat{\mathscr{m}}^r_x(2)$ and ${}_5\hat{\mathscr{m}}^r_x(3)$ for the rates of losing mother and for losing father respectively, estimated from the projections. For each estimate, we denote their standard errors with $\mathrm{SE}(\cdot)$.

We relate the two sets of estimates using latent variables in a hierarchical model. Equation 4 defines the latent variables and Equation 5 relates the latent variables based on their definitions.

$$
\begin{aligned}
{}_5\hat{p}^r_x(i) &\sim N(\phi^r_x(i), \mathrm{SE}({}_5\hat{p}^r_x(i))) \\
{}_5\hat{m}^r_x(u,v) &\sim N(\theta^r_x(u,v), \mathrm{SE}({}_5\hat{m}^r_x(u,v))) \\
{}_5\hat{\mathscr{p}}^r_x(l) &\sim N(\varphi^r_x(l), \mathrm{SE}({}_5\hat{\mathscr{p}}^r_x(l))) \\
{}_5\hat{\mathscr{m}}^r_x(l) &\sim N(\vartheta^r_x(l), \mathrm{SE}({}_5\hat{\mathscr{m}}^r_x(l))) \\
\\
&\text{for } i \in \{1,2,3,4\}, \\
&\quad (u,v) \in \{(1,2),(1,3),(1,4),(2,4),(3,4)\}, \\
&\quad l \in \{2,3\}, \\
&\quad x \in \{0,5,...,60\}, \text{ and} \\
&\quad r \in \{1,2,3\}
\end{aligned}
\tag{4}
$$

$$
\begin{aligned}
\varphi_x^r(2) &= \phi_x^r(2) + \phi_x^r(4) \\
\varphi_x^r(3) &= \phi_x^r(3) + \phi_x^r(4) \\
\vartheta_x^r(2) &= \frac{\left(\theta_x^r(1,2) + \theta_x^r(1,4)\right) \cdot \phi_x^r(1) + \theta_x^r(3,4) \cdot \phi_x^r(3)}{1 - \varphi_x^r(2)} \\
\vartheta_x^r(3) &= \frac{\left(\theta_x^r(1,3) + \theta_x^r(1,4)\right) \cdot \phi_x^r(1) + \theta_x^r(2,4) \cdot \phi_x^r(2)}{1 - \varphi_x^r(3)}
\end{aligned}
\tag{5}
$$

$$
\text{for } x \in \{0, 5, ..., 60\} \text{ and } r \in \{1, 2, 3\}
$$

We impose statistical models on the latent variables. For the proportion latent variables, $\phi_x^r(1)$, $\varphi_x^r(2)$, $\varphi_x^r(3)$, and $\phi_x^r(4)$, we assume a normal linear regression model on the logit of the variables as we expect the proportions to be monotonic in age x. We fit a regression model for the four variables with age x, racial and ethnic group r, and the state they represent as covariates. We also include all possible interactions in the regression model. We place non-informative $N(0, 1)$ priors on the coefficients and the intercept term, and Exp(1) on the standard deviation of the error term.

In the same manner, we impose a linear regression model on the log of the rate latent variables, $\theta_x^r(1, 2)$, $\theta_x^r(1, 3)$, $\theta_x^r(2, 4)$, and $\theta_x^r(3, 4)$. We again place non-informative $N(0, 1)$ on the coefficients and the intercept term. For the error term, we place Exp(5) prior on the standard deviation. For $\theta_x^r(1, 4)$, we place Gaussian random walk models on the log scale for each $r \in \{1, 2, 3\}$ independently. We use $\theta_0^r(1, 4) \sim \text{Exp}(20)$ as weakly informative priors on the rates at birth and place Exp(5) on the random walk standard deviation on the log scale.

3 Results

We fitted the Bayesian model in Method section using `stan`'s MCMC algorithm via its R interface (Team, 2024; Stan Development Team, 2024). Specifically, we used the library's NUTS algorithm over 4 chains with 5,000 sampling iterations each after 5,000 warmup iterations to draw posterior samples of the age-specific parent mortality estimates for the Hispanic, non-Hispanic black, and non-Hispanic white populations in the United States, 2020. We then used the posterior samples to construct the posterior distribution of the multistate life tables by parent mortality status.

Figure 2 displays the posterior medians and 95% credible intervals of life-table survivor counts. By construction, all groups begin with 100,000 survivors with both parents at birth and lose parents at different rates. The Hispanic population experiences the least parental losses over time with the highest number survivors with both parents alive by the age 65. On the other hand, the non-Hispanic white population has highest number of survivors who have lost both parents by 65. The non-Hispanic black survivors are most likely to have lost at least one parent from birth to 55. For all three groups, there are more survivors who have lost their fathers only than those who have lost their mothers only or both parents until they reach 55. In particular, the non-Hispanic black population displays higher risks of losing their fathers. There are visibly more non-Hispanic black survivors have lost their fathers until the age of 50 compared to the other two groups. There are also more black survivors younger than 30 who has lost their mothers while the difference is less pronounced.

Figure 3 displays the multistate life expectancy proportioned out of the total life expectancy above

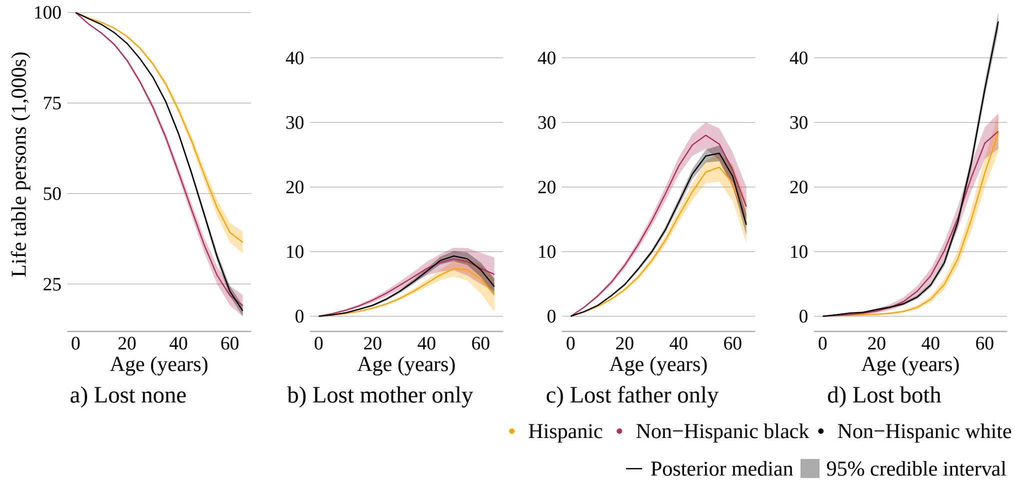

Figure 2. Posterior life-table survivor counts out of initial 100,000 by parent mortality state for the United States Hispanic, non-Hispanic black, and non-Hispanic white populations in 2020.

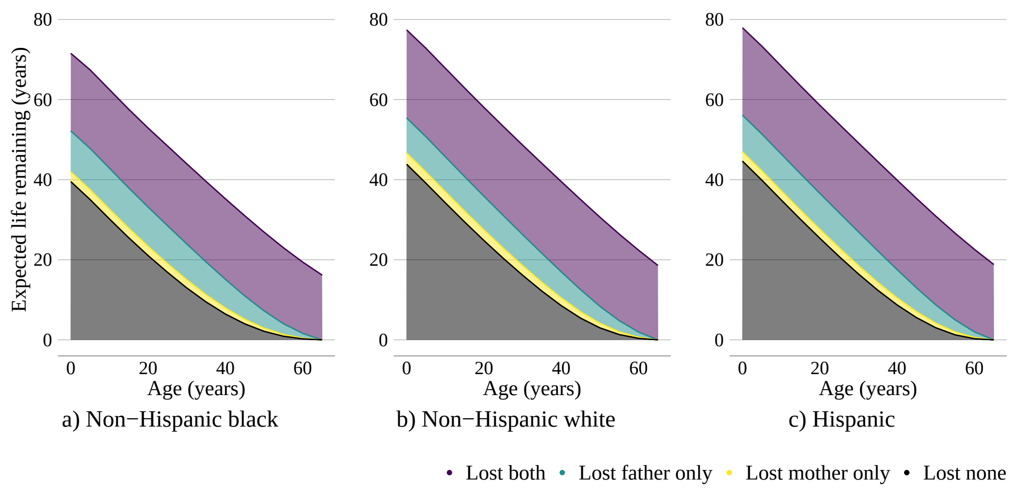

Figure 3. Posterior median life expectancy in each parent mortality state proportioned out of the total life expectancy above current age for the United State Hispanic, non-Hispanic black, and non-Hispanic white populations in 2020.

each age for the three groups. The height of shade for each state above each age represents the expected remaining number of years a person expects to live in the corresponding state. The non-Hispanic black population not only has the lowest life expectancy at birth but also the lowest expected number of years to live with both parents alive. On the other hand, the non-Hispanic black population expects to live with their mothers only for more years at birth than the other two groups despite shorter life expectancy in total. These results are direct implications of higher age-specific proportions of State (2) - Lost father only among the non-Hispanic black population - 0.005 (95%CI 0.004-0.006) at birth vs. 0.003 (95%CI 0.002-0.003) for the Hispanics and 0.002 (95% CI 0.002-0.003) for the. non-Hispanic whites. The expected remaining life with their mothers only remain higher for the non-Hispanic black population until they reach the age of 40.

4 Discussion

We constructed multistate life tables by parent mortality status using publicly available data. The result revealed notable racial discrepancies in the expected number of years to live with each parent for the non-Hispanic black population compared to the Hispanic and non-Hispanic white populations. Higher likelihoods of children and young adults living without their fathers combined with lower life expectancy overall mean that non-Hispanic black newborns are expected to live with their fathers alive for a shorter period. The current study's findings suggest that the non-Hispanic black population may be more vulnerable to risks associated with early parental loss.

The presented approach makes a few simplifying assumptions due to limitations in data availability. We assume that the proportions of person-years lived in each parent mortality state and parent mortality rates have remained consistent in the past 20 years. The assumption allowed us to utilize each 2021 SIPP respondent's parent mortality information beyond the reference year. Estimates based on the reference year parent mortality events only would lead to cells with smaller or no samples. To assess the level of bias introduced by the assumption, we plan to conduct a sensitivity analysis based on different ranges of retrospect information used for each respondent.

We also used the same estimates to proportion age-specific mortality rates by parent mortality status. This implies that mortality rates are constant between different those who have and those who have not experienced a parental loss. However, parental losses at earlier ages are associated with increased mortality risks (Niederkrotenthaler et al., 2012; Li et al., 2014; Enrico Debiasi and Eriksson, 2021). Currently, we lack the complete linked data between parent and offspring mortality needed to estimate the dependence.

This paper contributes to the understanding the systematic racial and ethnic disparities in mortality that are transferred over generations. We also propose a new framework to inform class demographic models. We combined existing classic demographic models with statistical models to generate probabilistic estimates informed by multiple sources of data.

References

Arias, E. and Xu, J. (2022), "United States LIfe Tables, 2020," *National Vital Statistics Report*, 71, https://www.cdc.gov/nchs/data/nvsr/nvsr71/nvsr71-01.pdf.

Caraballo, C., Massey, D. S., Ndumele, C. D., Haywood, T., Kaleem, S., King, T., Liu, Y., Lu, Y., Nunez-Smith, M., Taylor, H. A., Watson, K. E., Herrin, J., Yancy, C. W., Faust, J. S., and Krumholz, H. M. (2023), "Excess Mortality and Years of Potential Life Lost Among the Black Population in the US, 1999-2020," *JAMA*, 329, 1662–1670.

Caswell, H., Margolis, R., and Verdery, A. (2023), "The formal demography of kinship V: Kin loss, bereavement, and causes of death," *Demographic Research*, 49, 1163–1200.

Dwyer-Lindgren, L., Kendrick, P., Kelly, Y. O., Sylte, D. O., Schmidt, C., Blacker, B. F., Daoud, F., Abdi, A. A., Baumann, M., Mouhanna, F., Kahn, E., Hay, S. I., Mensah, G. A., Nápoles, A. M., Pérez-Stable, E. J., Shiels, M., Freedman, N., Arias, E., George, S. A., Murray, D. M., Phillips, J. W., Spittel, M. L., Murray, C. J., and Mokdad, A. H. (2022), "Life expectancy by county, race, and ethnicity in the USA, 2000–19: a systematic analysis of health disparities," *The Lancet*, 400, 25–38.

Enrico Debiasi, M. R.-F. and Eriksson, B. (2021), "The long-term consequences of parental death in childhood on mortality and the role of socioeconomic status: evidence from Sweden at the turn of the 20th century," *The History of the Family*, 26, 657–681.

Hummer, R. A. and Chinn, J. J. (2011), "RACE/ETHNICITY AND U.S. ADULT MORTALITY: Progress, Prospects, and New Analyses," *Du Bois Review: Social Science Research on Race*, 8, 5–24.

Li, J., Vestergaard, M., Cnattingius, S., Gissler, M., Bech, B. H., Obel, C., and Olsen, J. (2014), "Mortality after Parental Death in Childhood: A Nationwide Cohort Study from Three Nordic Countries," *PLOS Medicine*, 11, 1–13.

Niederkrotenthaler, T., Floderus, B., Alexanderson, K., Rasmussen, F., and Mittendorfer-Rutz, E. (2012), "Exposure to parental mortality and markers of morbidity, and the risks of attempted and completed suicide in offspring: an analysis of sensitive life periods," *Journal of Epidemiology & Community Health*, 66, 233–239.

Patterson, S. E., Verdery, A. M., and Daw, J. (2020), "Linked lives and childhood experience of family death on educational attainment," *Socius*, 6, 1–7, pMID: 34222657.

Schlüter, B.-S., Alburez-Gutierrez, D., Bibbins-Domingo, K., Alexander, M. J., and Kiang, M. V. (2024), "Youth Experiencing Parental Death Due to Drug Poisoning and Firearm Violence in the US, 1999-2020," *JAMA*, 331, 1741–1747.

Schoen, R. (1988), "Practical Uses of Multistate Population Models," *Annual Review of Sociology*, 14, 341–361, pMID: 12315442.

Stan Development Team (2024), "RStan: the R interface to Stan," R package version 2.32.6.

Stroebe, M., Schut, H., and Stroebe, W. (2007), "Health outcomes of bereavement," *The Lancet*, 370, 1960–1973.

Team, S. D. (2024), "Stan Modeling Language Users Guide and Reference Manual," .

Umberson, D., Olson, J. S., Crosnoe, R., Liu, H., Pudrovska, T., and Donnelly, R. (2017), "Death of family members as an overlooked source of racial disadvantage in the United States," *Proceedings of the National Academy of Sciences*, 114, 915–920.

United States Census Bereau (2023), "2021 Survey of Income and Program Participation Users' Guide," `https://www2.census.gov/programs-surveys/sipp/tech-documentation/methodology/2021_SIPP_Users_Guide_AUG22.pdf`.

United States Census Bureau (2023), "Survey of Income and Program Participation Datasets: 2021 SIPP Data," `https://www.census.gov/programs-surveys/sipp/data/datasets.2021.html`.