Survival Analysis: Impact of Marital Bereavement on Mortality among the Elderly

Yimin Chen (yc4195), Qingyue Zhuo (qz2493), Jiarui Yu (jy3360)

2023-12-11

Contents

Abstract

This study addresses the critical issue of understanding how marital bereavement affects mortality risks in the elderly. Previous studies have suggested that losing a spouse can significantly impact an individual's health, but quantifying this effect remains challenging. The primary objective of this project is to assess the impact of marital bereavement on the risk of death among the elderly, with a focus on how this risk is modified by demographic and health-related factors. We applied non-parametric and semi-parametric methods in the study. Our result shows that the change of bereavement status doesn't have a significant effect on mortality rate.

Background

The aging process is accompanied by various life events that can significantly impact health outcomes. The loss of a spouse stands out as a particularly traumatic experience with potential long-term effects on the survivor's well-being. Marital bereavement has been associated with increased mortality rates, commonly referred to as the "widowhood effect." Understanding the dynamics of how marital bereavement influences mortality is critical for public health and for developing interventions aimed at supporting the bereaved elderly population.

Objectives

The primary objective of this research is to quantify the impact of marital bereavement on all-cause mortality among the elderly. While previous studies have established a link between spousal loss and increased mortality risk, many questions remain regarding the mechanisms at play and the moderating role of demographic and health-related factors. This study seeks to examine the extent of these influences and captures relevant variables. Given the aging global population and the associated increase in the incidence of bereavement among older adults, this research holds significant value for health policy and geriatric care.

Exploratory Data Analysis

The study utilizes data from the Bereavement dataset (BRV) from the "biostat3" package in R. It includes 399 elderly individuals who were part of a larger longitudinal study. This dataset focuses on the elderly whose spouses were alive at the onset of the study, allowing for an in-depth analysis of the effects of marital bereavement on mortality.

The dataset includes a wide range of variables, including individual ID (id), couple identifiers (couple), dates of birth (dob), entry into the study (doe), exit from the study (dox), death of the spouse (dosp), and the mortality of the study subjects (fail). The primary outcome of the study is the time to event (death), with the event status indicated by the 'fail' variable, where 0 represents subjects who were alive at the end of the study (censored), and 1 indicates subjects who died during the study period (uncensored).

The data features right-censoring where the event of interest (death) has not occurred for all subjects by the end of the study period. The 'dosp' variable provides information about the date of the spouse's death; it is coded with a specific date if the spouse died during the study period or a uniform date (January 1, 2000) if the spouse was still alive at the study's conclusion.

Table 1: Summary Statistics for Date Variables

Variable	Description	Minimum Date	Maximum Date
dob	Date of Birth	1888-02-22	1906-03-12
doe	Date of Entry into Study	1981-01-15	1981-10-23
dox	Date of Exit from Study	1981-03-13	1991-01-01
dosp	Date of Death of Spouse	1981-05-22	2000-01-01

The study population primarily consists of elderly couples categorized into three distinct groups based on their age and living arrangements at the time of the initial interview. Group 1 comprises 226 index cases, all of whom were 75 years or older and lived in a two-person household. Group 2 includes 118 index cases with spouses younger than 75 years, also living in two-person households. Lastly, Group 3 contains 55 index cases who resided with a spouse and another person. The baseline characteristics of the sample include demographic information (age and sex), health status, and level of disability.

Table 2: Summary Statistics for Factor Variables

Variable	Description	Factor Top Counts	Meaning
group	Group	1: 597, 2: 302, 3: 137	1 = group 1, 2 = group 2, 3 = group 3
disab	Disability Level	0: 758, 1: 131, 3: 81, 2: 66	0 = none, 1 = mild, 2 = mod, 3 = severe
health	Health Status	2: 634, 1: 349, 0: 53	0 = poor, 1 = mod, 2 = good
sex	Sex	1: 554, 2: 482	1 = male, 2 = female
brv	Bereavement Status	0: 735, 1: 301	1 = bereaved, 0 = not bereaved
fail	Status at Exit	cen: 758, 1: 278	cen = censored, 1 = death
age	Age Band	(80: 408, (75: 326, (85: 258, (90: 38	(70,75], (75,80], ..., (95,100]

We performed a series of data transformations on the BRV dataset. Firstly, we created time variables `y_before_sp_dth` and `y_after_sp_dth` relative to the spouse's death. These represent the number of years before and after the spouse's death, calculated from the date of entry into the study (`doe`) and the date of the spouse's death (`dosp`), as well as the date of exit from the study (`dox`). Then, we split each subject's follow-up time into two records: one for the time before the spouse's death and another for the time after (if the spouse died during the study). The dataset is further modified to calculate the time each subject was at risk of the event (death) after the spouse's death. This is represented by the `t_sp_at_risk` variable. We created a new binary variable, `brv`, to indicate whether the subject was bereaved (1) or not (0) based on whether `y_after_sp_dth` is greater than 0. We also created additional time variables: `age_sp_dth` (age at spouse's death), `age_start` (age at the start of the observation period), and `age_end` (age at the end of the observation period). New variables `t_at_risk` (time at risk within each age band) and `age` (categorical age band) are created to reflect the duration and age category of each interval.

Table 3: Summary Statistics for Numeric Variables

Variable	Description	Mean	Median	Sd
y_before_sp_dth	Years Before Spouse's Death	-8.07	-5.58	7.89
y_after_sp_dth	Years After Spouse's Death	-2.72	0.00	7.13
t_sp_at_risk	Time at Risk After Spouse's Death	5.35	5.25	3.03
age_sp_dth	Age At Spouse's Death	87.71	85.80	7.26
age_start	Start Age	81.33	80.00	3.80
age_end	End Age	83.68	84.81	3.69
t_at_risk	Time at risk	2.35	2.06	1.61

By the end of this process, the dataset is transformed to facilitate survival analysis by accommodating the change in risk status due to bereavement and by considering the age of the subjects as they progress through different age bands. This detailed partitioning allows for a nuanced analysis of how the risk of death changes with age and bereavement status over time.

Methods

Life Table

To investigate our research questions, we first applied the Life-table, which was constructed using standard life table methodology.

Probability function at t_middle

$$t_{mi} = \frac{t_i + t_{i-1}}{2}$$

$$f(t_{mi}) = \frac{\hat{S}_L(t_{i-1}) - \hat{S}_L(t_i)}{t_i - t_{i-1}}$$

$$\hat{S}_L(t_{mi}) = \frac{\hat{S}_L(t_{i-1}) + \hat{S}_L(t_i)}{2}$$

Number of events per person-time-units

$$\hat{h}(t_{mi}) = \frac{d_i}{I(t_i - t_{i-1})(n_i' - \frac{d_i}{2})}$$

Based on the definition

$$\hat{h}(t_{mi}) = \frac{f(t_{mi})}{S(t_{mi})} = \frac{2f(t_{mi})}{S(t_i) + S(t_{i-1})}$$

Variance

$$\text{var}\{\hat{h}(t_{mi})\} = \frac{(\hat{h}(t_{mi}))^2}{n_i' q_i} \left(1 - \frac{(\hat{h}(t_{mi})(t_i - t_{i-1}))^2}{2}\right)$$

Kaplan-Meier and Fleming-Harrington model

For nonparametric estimator, Kaplan-Meier(KM) model and Fleming-Harrington(FH) model were used to measure the fraction of subjects living for a certain amount of time after the change in bereavement status.

The Kaplan-Meier estimator

$$\hat{S}_K(t) = \begin{cases} 1 & \text{if } t < t_1 \\ \prod_{t_i \leq t}[1 - \frac{d_i}{n_i}] & \text{if } t \geq t_1 \end{cases}$$

note: $d_i = \# \ of \ failure \ at \ time \ t_i$, $n_i = \# \ at \ risk \ at \ t_i^-$, $c_i = \# \ censored \ during \ the \ interval \ [t_i, t_{i+1}]$

The Fleming-Harrington estimator

$$\hat{S}_F(t) = \begin{cases} 1 & \text{if } t < t_1 \\ \prod_{t_i \leq t} exp[-\frac{d_i}{n_i}] & \text{if } t \geq t_1 \end{cases}$$

Log-rank Test

We then use a non-parametric log-rank test, which makes no distributional assumptions about our data. The log-rank test compares differences between expected and observed events at each event time point, k, to derive the following test statistics:

The likelihood function is defined as:

$$L = \sum_{i=1}^{k}(d_{0i} - e_{0i});$$

And the variance is given by:

$$\text{var}(L) = \sum_{i=1}^{k}\frac{n_{0i}n_{1i}d_i(n_i - d_i)}{n_i^2(n_i - 1)};$$

A significant result from the log-rank test indicates a statistically significant difference between the two groups. Notably, the log-rank test does not allow us to adjust for covariates or prognostic variables.

Cox Proportional Hazard Model

As a further step in this analysis, we consider the Cox Proportional Hazard (Cox PH) model, which allows us to model the hazard ratio based on covariates to understand their impact on the survival function.

The Cox PH typically takes the form:

$$h(t \mid Z = z) = h_0(t)e^{\beta' z}$$

$$S(t|Z = z) = e^{-\int h_0(t)e^{\beta z}dt} = S_0(t)^{e^{\beta z}}$$

After the log-log transformation

$$log\{-logS(t|Z = z)\}$$

we will have

$$log\{-log\hat{S}(t|Z = z)\} - log\{-log\hat{S}_0(t)\} = \beta$$

Results

Non-parametric Estimate

Lifetable

The life table analysis reveals a gender-based discrepancy in survival times, indicating that the median survival time for females falls between 1800 and 2100 days, whereas for males, it ranges from 1500 to 1800 days. Notably, females exhibit a significantly longer median survival duration compared to their male counterparts. That is also true for other time points.

Life Table of Male Observations

	tstart	tstop	nsubs	nlost	nrisk	nevent	surv	pdf	hazard	se.surv	se.pdf	se.hazard
0-300	0	300	260	1	259.5	25	1.0000000	0.0003211	0.0003374	0.0000000	6.11e-05	0.0000674
300-600	300	600	234	3	232.5	25	0.9036609	0.0003239	0.0003788	0.0183162	6.15e-05	0.0000756
600-900	600	900	206	4	204.0	29	0.8064930	0.0003822	0.0005101	0.0245819	6.68e-05	0.0000944
900-1200	900	1200	173	3	171.5	26	0.6918445	0.0003496	0.0005468	0.0288704	6.48e-05	0.0001069
1200-1500	1200	1500	144	7	140.5	22	0.5869585	0.0003064	0.0005663	0.0309662	6.21e-05	0.0001203
1500-1800	1500	1800	115	8	111.0	29	0.4950504	0.0004311	0.0010017	0.0317168	7.41e-05	0.0001839
1800-2100	1800	2100	78	9	73.5	26	0.3657129	0.0004312	0.0014325	0.0312268	7.73e-05	0.0002744
2100-2400	2100	2400	43	14	36.0	26	0.2363451	0.0005690	0.0037681	0.0286923	9.07e-05	0.0006096
2400-2700	2400	2700	3	1	2.5	16	0.0656514	0.0014006	-0.0096970	0.0193600	NaN	NaN

Figure 1: Lifetable (male)

Life Table of Female Observations

	tstart	tstop	nsubs	nlost	nrisk	nevent	surv	pdf	hazard	se.surv	se.pdf	se.hazard
0-300	0	300	295	1	294.5	25	1.0000000	0.0002830	0.0002955	0.0000000	5.41e-05	0.0000590
300-600	300	600	269	3	267.5	25	0.9151104	0.0002851	0.0003268	0.0162413	5.45e-05	0.0000653
600-900	600	900	241	4	239.0	29	0.8295860	0.0003355	0.0004306	0.0219548	5.91e-05	0.0000798
900-1200	900	1200	208	3	206.5	26	0.7289250	0.0003059	0.0004479	0.0260603	5.71e-05	0.0000876
1200-1500	1200	1500	179	7	175.5	22	0.6371475	0.0002662	0.0004458	0.0283207	5.44e-05	0.0000948
1500-1800	1500	1800	150	8	146.0	29	0.5572771	0.0003690	0.0007351	0.0294482	6.44e-05	0.0001357
1800-2100	1800	2100	113	9	108.5	26	0.4465851	0.0003567	0.0009075	0.0299248	6.55e-05	0.0001763
2100-2400	2100	2400	78	14	71.0	26	0.3395693	0.0004145	0.0014943	0.0292004	7.39e-05	0.0002856
2400-2700	2400	2700	38	1	37.5	16	0.2152200	0.0003061	0.0018079	0.0268227	6.94e-05	0.0004350
2700-3000	2700	3000	21	13	14.5	19	0.1233928	0.0005390	0.0126667	0.0232088	7.44e-05	NaN

Figure 2: Lifetable (female)
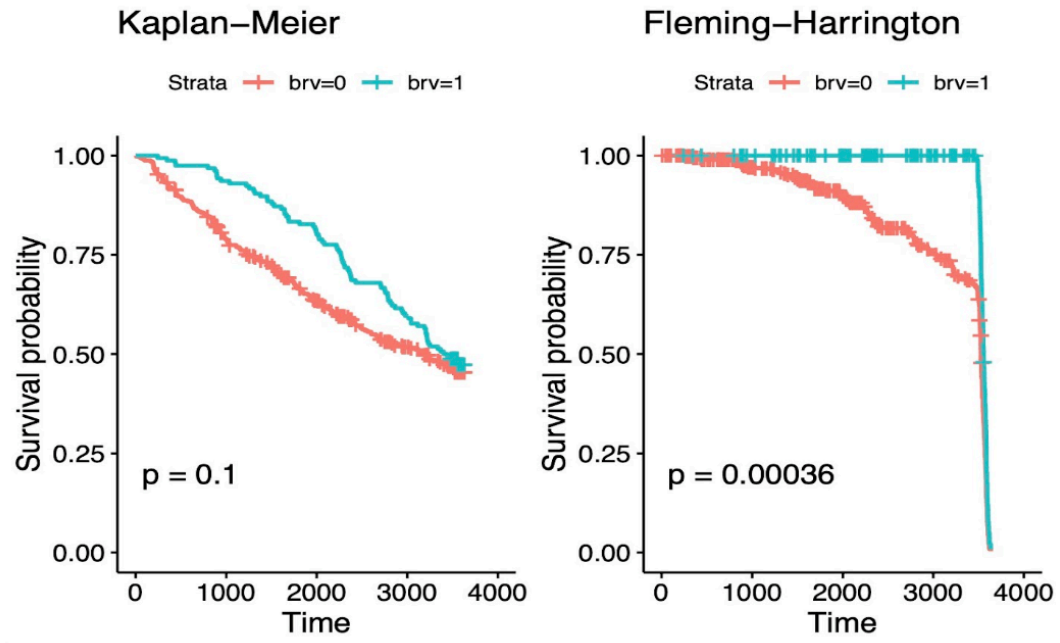
The Kaplan-Meier and Fleming-Harrington model

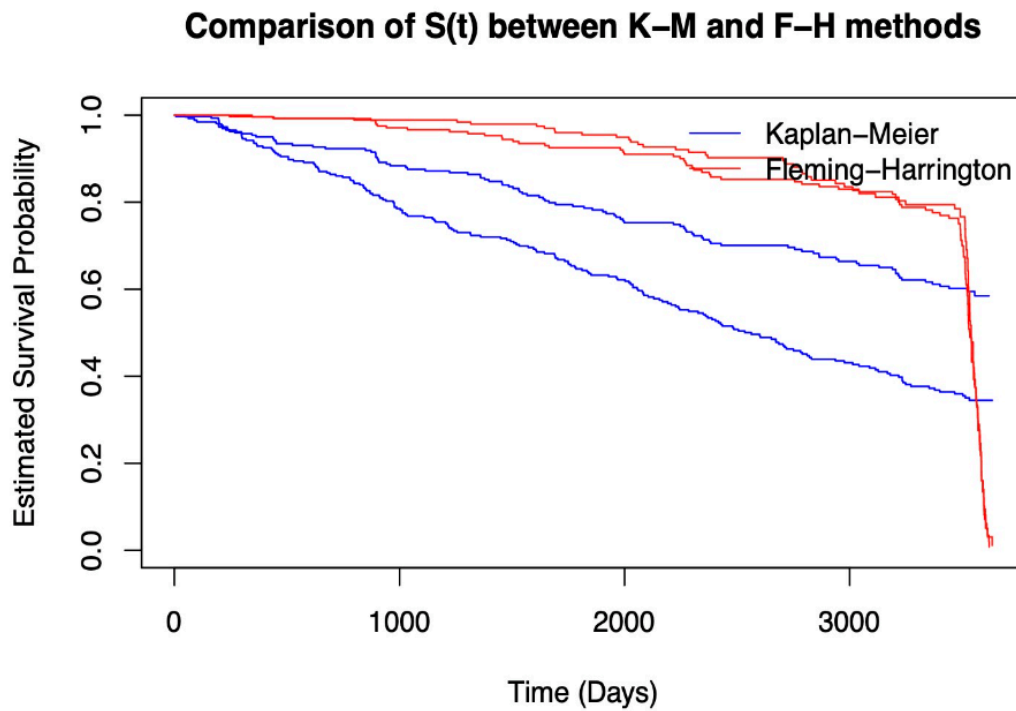
Figure 3: Kaplan-Meier modeland Fleming-Harrington model

Figure 4: Comparison of Kaplan-Meier modeland Fleming-Harrington model

In the investigation of the impact of marital bereavement status on survival time, both the Kaplan-Meier (KM) Estimator and Fleming-Harrington (FH) Estimator were employed. The KM estimator yielded a p-value of 0.01, surpassing the conventional significance threshold of 0.05. This result implies that the observed change in marital bereavement status is not statistically significant in relation to the survival time among the elderly population. Conversely, the FH estimator exhibited a p-value below 0.05, suggesting statistical significance and implying that the change in bereavement status has a notable impact on survival time according to this estimator. Despite similar trends between the KM and FH models, there is an observed difference in the estimation value across time. the FH estimator indicates a higher value compared to the KM estimator most of the time. However, towards the conclusion of the study period, there is a notable and abrupt decrease in the FH estimator.

Hypothesis Testing

Log-Rank test

In our study examining the potential influence of sex, group membership, and bereavement status on the mortality of elderly individuals, log-rank tests were employed. The obtained p-values reveal distinctive patterns:

Sex: The p-value associated with sex is 3e-08, which is exceedingly low. This extremely small p-value indicates a highly statistically significant difference in survival between males and females. The findings suggest that the impact of sex on survival among the elderly is substantial and merits further investigation.

Group: The p-value linked to the group category is 0.4. This relatively high p-value suggests the absence of a statistically significant difference in survival across the various groups. The evidence does not support the hypothesis that group membership significantly influences the mortality of elderly individuals. Further

analysis and consideration of potential covariates may be necessary to confirm these results.

Bereavement Status: The p-value associated with changes in bereavement status is 0.1. This p-value indicates that the observed difference in survival between the bereaved and non-bereaved groups is not statistically significant. Therefore, based on our analysis, there is insufficient evidence to conclude that changes in bereavement status significantly impact the survival of elderly individuals. However, caution should be exercised, and additional investigations may be warranted to explore potential nuances or interactions within the data. In summary, our log-rank tests reveal a robust and statistically significant association between sex and survival, while group membership and changes in bereavement status do not exhibit significant impacts on mortality among the elderly population in our study.

Semi-parametric Model (PH model)

In our investigation, the Cox proportional hazards model was employed to assess the impact of bereavement, and these estimates were juxtaposed against those derived from Poisson regression. The resulting outcome table highlights noteworthy findings:

Significant Predictors: Sex, disability level, and health status demonstrated p-values smaller than 0.05 in the Cox model, signifying their statistical significance in predicting patient mortality. These variables emerged as significant predictors of the outcome within the framework of the Cox proportional hazards model.

Stepwise Selection: Employing stepwise selection, a refined model was constructed, encompassing three variables: sex, disability level, and health status. The stepwise selection process aims to identify the most influential predictors while optimizing model fit.

Final Model Evaluation: The final model, comprising the aforementioned variables, was assessed using the Akaike Information Criterion (AIC). Notably, this model yielded the lowest AIC value, standing at 2702.01. The minimized AIC underscores the model's improved goodness of fit and parsimony, suggesting its efficacy in capturing the essential predictors contributing to the mortality of patients under consideration.

	coef	exp(coef)	se(coef)	z	Pr(>|z|)
brv1	0.003182	1.003187	0.143946	0.022106	0.982363
sex2	-0.520998	0.593927	0.132939	-3.919082	0.000089
group	0.026896	1.027261	0.084823	0.317085	0.751179
disab	0.254538	1.289866	0.057233	4.447410	0.000009
health	-0.460763	0.630802	0.101012	-4.561458	0.000005

Figure 5: Stepwise Selectioin Model Results

Model Checking

The Cox proportional hazards (PH) model makes two major assumptions. One of them is that the hazard functions for the survival curves of different strata will be proportional over the period of time t, and the other one is the relationship between the log of hazard $h(t)$ and each of the covariates will be linear. The following introduces some graphical methods and residuals plot for assumption checking.

Graphical Approach One of strategies for PH assumption checking is to compare the survival curves visually. Mentioned in the statistical methods part above, $S(t|Z = z) = e^{-\int h_0(t)e^{\beta z}dt} = S_0(t)^{e^{\beta z}}$. After the log-log transformation (i.e., $log\{-logS(t|Z = z)\}$), we will have $log\{-log\hat{S}(t|Z = z)\} - log\{-log\hat{S}_0(t)\} = \beta$, If the two curves are relatively subjective, the proportional assumption predicts that they will be approximately paralleled. The transformed survival functions calculated using the K-M estimator are shown below, along with the log of time (days).
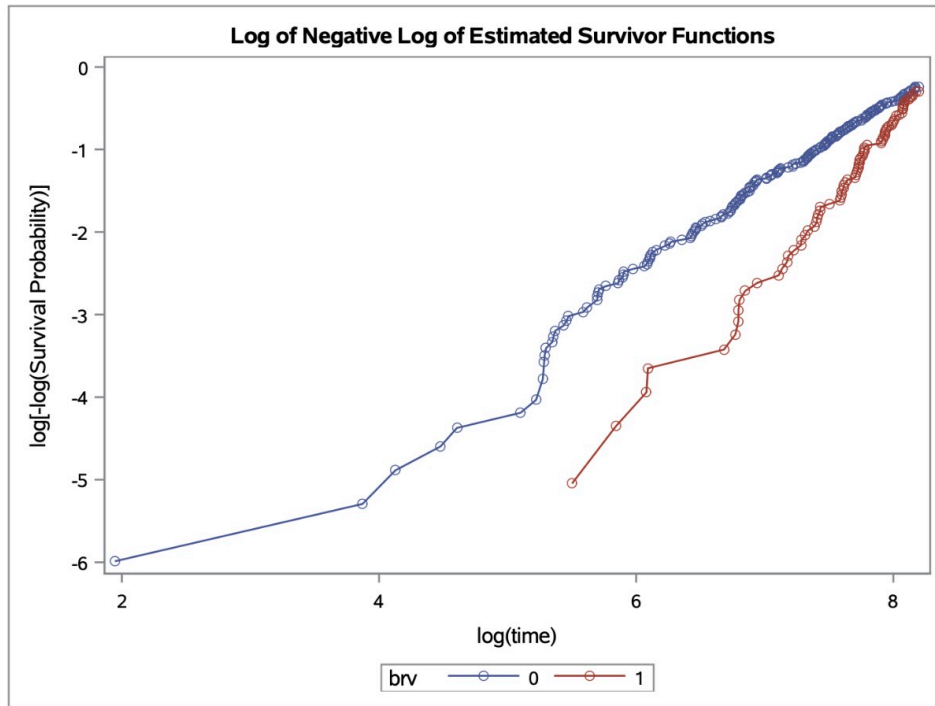
Figure 6: Log of Negative Log of Estimated Survival Functions

Residuals Test Two steps are involved in the evaluation of PH assumptions. Begin by assessing the significance of each covariate using a global test (e.g., Score test, partial likelihood ratio test), and then validate the results using the Schoenfeld residuals scatter plot. For these techniques, we examine all three important variables produced from the model above: gender, disability level, and health status. The outcomes are presented below.
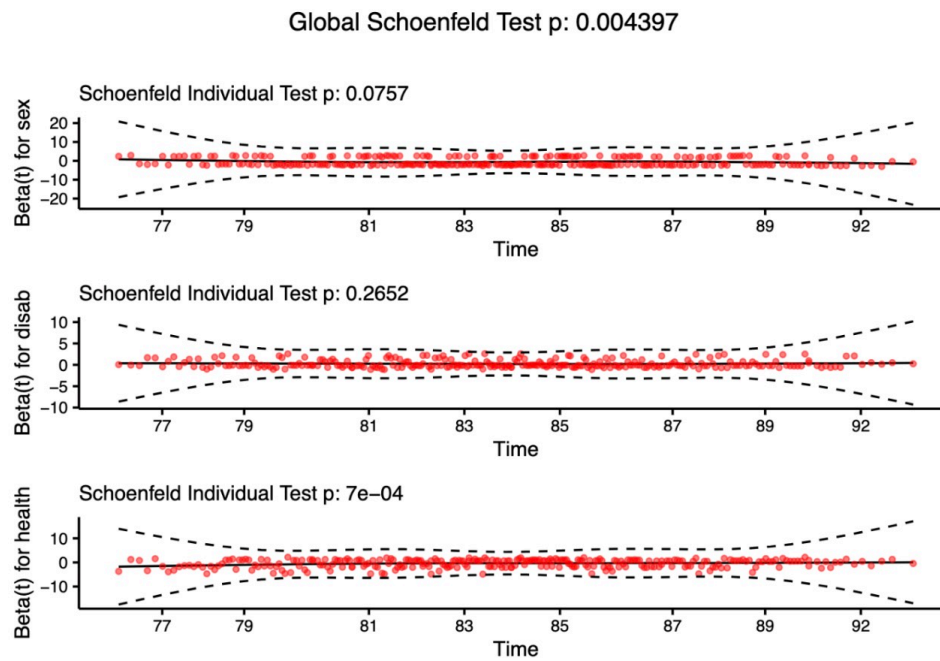
Figure 7: The Global Schoenfeld Test

The Global Schoenfeld Test has a p-value of 0.004397, which is below the common alpha level of 0.05, indicating that there is statistically significant evidence against the proportional hazards assumption across all covariates in the model. This suggests that the hazard ratios for the covariates in the model are not consistent over time. The individual Schoenfeld Test for the 'sex' variable has a p-value of 0.0757, which is above the 0.05 threshold, suggesting that there is no significant evidence against the proportional hazards assumption for the 'sex' variable. The plot for 'sex' shows residuals scattered around the zero line without any apparent trend, which is in line with the proportional hazards assumption. The individual Schoenfeld Test for 'disab' (presumably 'disability') has a p-value of 0.2652, also indicating no significant violation of the proportional hazards assumption for this variable. Lastly, the individual Schoenfeld Test for 'health' has a very low p-value of 7e-04, indicating a significant violation of the proportional hazards assumption for this variable. In summary, the Cox model may not be appropriate for the 'health' variable due to violation of the proportional hazards assumption. The 'sex' and 'disab' variables do not show evidence against this assumption. It may be necessary to explore time-varying covariates or alternative models for the 'health' variable to adequately model these data.
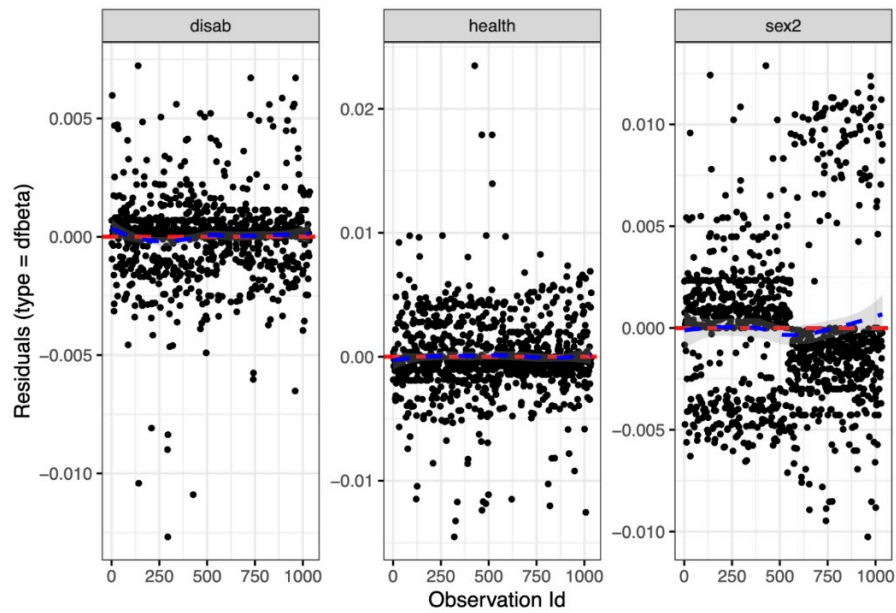
Figure 8: Influential Outliers

Also, from the plot above , it seems that for all covariates, there is no clear pattern of systematic influence, and no single observation appears to be particularly influential, as most of the dfbeta values are close to zero and within the confidence band.

Discussion

The primary limitation of our study pertains to the modest size of our dataset, comprising 399 observations. This smaller sample size introduces potential challenges, including Reduced Statistical Power, Increased Type II Error Risk, Unreliable Estimates, Limited Generalizability, and Instability in Model Fitting. These limitations may impact the robustness and generalizability of our findings. Additionally, another noteworthy limitation stems from the absence of a "treatment" variable in our dataset. Consequently, our analysis lacks insight into potential health interventions or treatments related to the observed outcomes. Instead, our focus centers on investigating the impact of bereavement on the life expectancy of the elderly. To address this limitation and enhance the scope of our study, we propose the inclusion of additional datasets capturing information on interventions relevant to bereavement, such as counseling, psychotherapy, or physical exercise provided to older individuals. This augmentation would enable us to assess the influence of these interventions on survival outcomes, thereby enriching the depth and applicability of our analysis.

Conclusion

References

[1] López, P., Rodríguez, A. C., Escapa, S. (2022). "Psychosocial effects of gentrification on elderly people in Barcelona from the perspective of bereavement." Emotion, Space and Society, 43, 100880.

[2] Matthys, O., Dierickx, S., Deliens, L., Lapeire, L., Hudson, P., Van Audenhove, C., De Vleminck, A., Cohen, J. (2023). "Is pre-bereavement collaboration between family caregivers and healthcare professionals associated with post-bereavement emotional well-being? A population-based survey." Patient Education and Counseling, 110, 107654.

[3] Denckla, C. A., Hahn, J., Cowden, R. G., Ho, S., Gao, K., Espinosa Dice, A. L., Jha, S. C., Kang, J. H., Shear, M. K. (2023). "Bereavement, Memorial Attendance, and Mental Health During the COVID-19 Pandemic: Longitudinal Results from the Nurses' Health Study." The American Journal of Geriatric Psychiatry, 31(12), 1045-1057.

Appendix

For codes please click *here*.