Survival Analysis on Spousal Mortality

A Project report submitted in partialfulfillment of requirements for the Degree of M.Sc. (Statistics) 3^{rd} Semester

With specialization in Industrial Statistics

DEPARTMENT OF STATISTICS
SCHOOL OF MATHEMATICAL SCIENCES
KBC NORTH MAHARASHTRA UNIVERSITY
JALGAON - 425001

2024-2025

By

Ms. Borse Harshada Sudam (Seat no: 366341)

Mr. Patil Vivek Vilasrao (Seat no: 366358)

Mr. Patil Yogesh Yuvraj (Seat no: 366360)

Under Guidance of:

Prof. R. L. Shinde

Senior Professor, Department of Statistics

<u>CERTIFICATE</u>

This is to certify that **Ms. Borse Harshada Sudam, Mr. Patil Vivek Vilasrao and Mr. Patil Yogesh Yuvraj** students of M.Sc. (Statistics) with specialization in Industrial Statistics, at Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon have successfully completed their project work entitled **“Survival Analysis on Spousal Mortality”** based on the Primary data collected and review of two research papers a part of M.Sc. (Statistics) program under my guidance and supervision during the academic year 2024-2025.(Sem-III)

(Prof. R. L. Shinde)
Project Guide

ACKNOWLEDGEMENT

On this note, firstly we want to convey our regards to Prof. R. L. Shinde, Department of Statistics, Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon for giving his guidance during this project. It was a great pleasure to work under him to make our work comfortable to handle.We also thankful to Dr. R. D. Koshti for his support during project.

It would be unfair to go without acknowledging our other teaching staff and beloved friends at from the Department of Statistics KBCNMU, Jalgaon for their precious support to us.

Ms. Borse Harshada Sudam

Mr. Patil Vivek Vilasrao

Mr. Patil Yogesh Yuvraj

INDEX

Sr. No.	Chapter Title	Pg. No.
1.	**Introduction**	**1**
	1.1 Abstract	
	1.2 Motivation	
	1.3 Objectives	
2.	**Data Descriptions and Data Preprocessing**	**3**
	2.1 Data Sources	
	2.2 Data Collection	
	2.3 Data Descriptions	
	2.4 Data Variables	
	2.5 Data Preprocessing	
3.	**Methodology**	**8**
	3.1 Non-Parametric Models	
	3.1.1 Censoring	
	3.1.2 Survival function and Hazard function	
	3.1.3 Kaplan-Meier Estimator	
	3.2 Cox Proportional Hazards Model	
	3.3 Log-rank Test	
	3.4 Copula Model Approach	
4.	**Data Analysis**	**16**
	4.1 Survival Analysis on observed life of spouses	
	4.2 Survival Analysis of remaining life of spouse after the death of other spouse	
	4.3 Log rank & Copula test.	
5.	**Review of research paper**	**28**
	5.1 Review of "The Linked Lives, Dependent Demise? Survival Analysis of Husbands and Wives."	
	5.2 Review of "The Effect of Widowhood on Mortality by the Causes of Death of Both Spouses."	
	Conclusions	**37**
	Appendix	**38**
	References	**45**

CHAPTER 1: INTRODUCTION

Survival analysis is branch of statistics in which study of distribution of life length is studied. Here the random variable studied is "Time to event occurred". Here the goal is to analyze and model the data where the outcome is the time until the occurrence of an event of interest. One of the main challenges in this context is the presence of instances whose event outcomes become unobservable after a certain time point or when some instances do not experience any event during the monitoring period. Such Phenomenon is called censoring which can be effectively handled using survival analysis techniques.

This project investigates the survival times of paired individuals (husband and wife) to understand how the death of one partner influences the survival of the other. Using primary data, we focus on the **age of the husband** at the time of death, whether his death is censored or observed, and similarly, the **age of the wife** at the time of death along with her death status. The goal is to explore the dependency between the survival times of the two partners and determine whether the death of one partner affects the remaining life of the other.

To analyse this relationship, we apply **non-parametric methods** such as the **Kaplan-Meier estimator**, which allows us to estimate survival probabilities without making assumptions about the data's distribution. This provides flexibility in model survival trends. Additionally, **copula models and cox model** are employed to capture the dependence structure between the survival times of the husband and wife. These models help assess whether the survival of one partner is influenced by the other. The combination of non-parametric methods and copula models offers a comprehensive framework for understanding paired survival dynamics and provides insights into the effects of one partner's death on the other's remaining life. In addition, we have read papers that discuss spousal dependency in the mortality of partners. These studies provide valuable insights into how the death of one spouse can significantly impact the survival probability of the other, highlighting the interdependence of spousal mortality.

Motivation

The death of a spouse can have profound emotional, psychological, and physiological consequences on the surviving partner. Few studies have shown that the loss of a spouse often leads to increased stress, weakened immune systems, and adverse mental health outcomes, potentially resulting in a shortened life expectancy (Smith and Zick (2006)) and (Elwert and Christakis (2008)). While traditional survival analysis methods have been extensively used to study individual life expectancy and survival probabilities, there has been comparatively less focus on the interdependence between spousal deaths. This project is motivated by the need to explore the correlation between the survival times of spouses, particularly investigating how the death of one partner may influence the survival of the other.

By analysing local data on survival times of husbands and wives, the project aims to uncover the underlying factors that contribute to the dependency between spousal death events. The combination of parametric and copula models offers a novel approach to understanding spousal death dependency. By integrating these methodologies, the project seeks to provide valuable insights into the dynamics of spousal survival, contributing to a deeper understanding of the emotional, social, and psychological factors that affect life expectancy in couples. This research will have broader applications in the fields of public health, psychology, and social sciences, providing a fresh perspective on survival analysis and the dependency of one partner's survival on the other.

Objectives:

The primary objective of this project is to examine the relationship between the survival times of spouses and the impact of one spouse's death on the life expectancy of the surviving partner. The specific objectives of the project are:

1. **Primary Data Collection and Analysis**: Collect and analyse data on the death records of husbands and wives to explore the survival probabilities of the surviving spouse. We will apply **non-parametric** survival analysis methods (e.g., Kaplan-Meier estimator) to analyse the data.
2. **Use of Copula Models and Cox Model to Assess Dependency**: Employ **copula models** to investigate the dependency between the survival times of husbands and wives. The goal is to understand whether the death of one partner affects the life expectancy of the other partner, highlighting any correlation in their survival patterns.
3. **Review of Related Research**: Conduct a comprehensive review of existing research papers related to **husband and wife mortality** and the dependency between their survival times. This review will provide a foundation for understanding current trends in the field and help compare the findings of this project with prior research.

Chapter 2: Data and Data Descriptions

This chapter outlines the key steps taken to prepare the dataset for analysis. It begins with data collection and cleaning to identify and correct any inconsistencies in birth years, ages, and survival statuses. Next, the survival time for each individual is calculated by determining the difference between age (or age at death) and birth year. The chapter also discusses handling censored data, ensuring accurate survival function estimation. Finally, several new variables are derived, such as the complete life expectancy of both spouses and the remaining age of each spouse after the other's death. These preprocessing steps lay the foundation for conducting a comprehensive survival analysis.

2.1 Data Sources:

Data Collected on 350 married couples from local community. This data consists of survival times of husbands and wives. The primary goal is to investigate whether the death of one spouse has a measurable impact on the life expectancy of the surviving spouse. This dataset includes variables such as the death date, age at death, gender, and survival status of the surviving partner.

2.2 Data Collection:

To gather relevant data, we designed a straightforward questionnaire aimed at collecting primary information directly from families in the community or nearby areas. This method ensured that the data was specific, reliable, and suited for the objectives of the project.

Questionnaire Design:

The questionnaire was developed with simplicity in mind to encourage participation and accurate responses. It comprised the following key questions:

1. Do you know any married couple in which either one or both partner is (are) passed away in your family or nearby? Respondents were asked to indicate "Yes" or "No." If the response was "Yes," they proceeded to answer the following questions:

    i) Birth year of the husband.

    ii) Current status of husband's survival (1: Alive, 0: Dead)

    iii) Birth year of the wife.

    iv) Current status of wife's survival (1: Alive, 0: Dead)

    v) Age at death of the husband (if died).

    vi) Age at death of the wife (if died).

2.3 Data Descriptions:

2.3.1 Data Variables

The primary data collected for this project focuses on the birth year, age (current or at death) and survival status of husbands and wives in a local community. These variables are essential for investigating whether the death of one spouse impacts on the remaining life of another spouse. The specific variables assigned to the collected data are given below:

1. **Birth Year of Husband (BYH):** The year the husband was born.
2. **Birth Year of Wife (BYW):** The year the wife was born.
3. **Age of Husband (X1):** Age of the husband at the time of death, or if alive, his current age.
4. **Censored of Husband (SX1):** Status of the husband (1 if died, 0 if alive).
5. **Age of Wife(X2):** Age of the wife at the time of death, or if alive, her current age.
6. **SX2:** Status of the wife (1 died, 0 if alive).

Our Collected Data as below:

Sr. No.	BYH	BYW	X1	SX1	X2	SX2		Sr. No.	BYH	BYW	X1	SX1	X2	SX2
1	1933	1940	74	1	76	1		176	1953	1960	59	1	64	0
2	1945	1954	60	1	70	0		177	1940	1946	81	1	70	1
3	1934	1941	47	1	83	0		178	1929	1936	76	1	85	1
4	1930	1935	66	1	59	1		179	1955	1959	69	0	64	1
5	1953	1960	52	1	60	1		180	1952	1958	72	1	66	0
6	1946	1952	76	1	72	0		181	1940	1945	81	1	79	0
7	1970	1975	45	1	49	0		182	1952	1962	70	1	62	0
8	1951	1953	56	1	71	0		183	1958	1964	66	0	56	1
9	1947	1955	76	1	65	1		184	1967	1977	54	1	47	0
10	1922	1925	60	1	70	1		185	1960	1967	64	0	55	1
11	1936	1942	64	1	82	0		186	1934	1941	76	1	64	1
12	1948	1955	64	1	55	1		187	1953	1961	62	1	58	1
13	1940	1945	75	1	79	0		188	1927	1936	80	1	74	1
14	1945	1950	70	1	48	1		189	1949	1956	64	1	63	1
15	1960	1968	59	1	56	0		190	1967	1976	57	0	43	1
16	1938	1948	68	1	72	1		191	1966	1971	54	1	53	0
17	1919	1925	78	1	64	1		192	1971	1976	53	0	46	1
18	1940	1945	73	1	79	0		193	1948	1956	76	0	64	1
19	1976	1980	31	1	44	0		194	1960	1967	62	1	57	0
20	1960	1966	42	1	58	0		195	1951	1959	73	0	63	1
21	1940	1949	33	1	75	0		196	1963	1970	45	1	54	0
22	1972	1976	50	1	48	0		197	1957	1965	67	0	57	1
23	1955	1960	55	1	64	0		198	1930	1934	85	1	90	0
24	1968	1974	56	0	40	1		199	1929	1936	77	1	85	1
25	1960	1965	64	1	59	0		200	1950	1960	57	1	72	0
26	1943	1950	76	1	74	0		201	1966	1971	55	1	53	0
27	1944	1951	70	1	52	1		202	1973	1979	50	1	45	0
28	1944	1952	42	1	60	1		203	1952	1960	66	1	64	0
29	1964	1971	58	1	49	1		204	1964	1968	60	0	50	1
30	1975	1983	44	1	41	0		205	1930	1938	70	1	68	1
31	1955	1962	61	1	62	1		206	1970	1975	50	0	43	1
32	1945	1952	70	1	72	0		207	1940	1946	82	1	69	1
33	1946	1953	57	1	62	1		208	1961	1969	60	1	53	1
34	1968	1977	52	1	47	0		209	1970	1977	54	0	45	1
35	1959	1966	55	1	50	1		210	1971	1980	51	1	44	0
36	1936	1939	76	1	85	0		211	1953	1958	71	0	64	1
37	1931	1935	80	1	85	1		212	1941	1951	81	1	70	1
38	1964	1971	47	1	53	0		213	1945	1951	65	1	71	1

Sr. No.	BYH	BYW	X1	SX1	X2	SX2		Sr. No.	BYH	BYW	X1	SX1	X2	SX2
39	1969	1978	50	1	46	0		214	1961	1971	63	0	53	1
40	1936	1941	88	0	79	1		215	1957	1963	67	0	59	1
41	1948	1953	48	1	51	1		216	1928	1936	70	1	81	1
42	1950	1960	57	1	64	0		217	1946	1956	60	1	68	1
43	1967	1973	38	1	44	1		218	1968	1972	54	1	52	0
44	1965	1975	59	0	47	1		219	1954	1961	70	0	58	1
45	1964	1971	56	1	53	0		220	1961	1969	53	1	55	0
46	1971	1978	47	1	46	0		221	1952	1958	55	1	64	1
47	1967	1973	56	1	31	1		222	1951	1961	71	1	57	1
48	1973	1980	51	0	40	1		223	1951	1957	66	1	67	0
49	1943	1953	70	1	71	0		224	1946	1955	65	1	53	1
50	1946	1953	78	1	71	0		225	1940	1946	79	1	69	1
51	1943	1948	81	0	60	1		226	1946	1953	72	1	71	1
52	1930	1935	83	1	89	0		227	1963	1968	50	1	56	0
53	1934	1939	68	1	85	0		228	1959	1969	60	1	55	0
54	1954	1962	50	1	61	1		229	1938	1947	75	1	77	0
55	1938	1946	77	1	57	1		230	1941	1948	83	0	67	1
56	1941	1949	60	1	47	1		231	1959	1968	49	1	58	0
57	1966	1973	58	1	51	0		232	1958	1967	66	0	55	1
58	1944	1951	80	0	49	1		233	1942	1952	80	1	53	1
59	1942	1949	82	0	51	1		234	1971	1974	53	1	50	0
60	1967	1977	57	0	45	1		235	1942	1949	68	1	65	1
61	1945	1950	63	1	64	1		236	1952	1960	49	1	56	1
62	1943	1950	81	0	70	1		237	1936	1942	83	1	78	1
63	1930	1936	66	1	72	1		238	1963	1969	40	1	55	1
64	1944	1950	66	1	74	0		239	1965	1970	60	0	54	1
65	1940	1949	81	1	70	1		240	1967	1976	93	0	37	1
66	1944	1952	80	0	71	1		241	1952	1959	72	0	60	1
67	1932	1942	78	1	82	0		242	1957	1964	58	1	60	0
68	1937	1945	81	1	75	1		243	1941	1947	65	1	70	1
69	1966	1973	42	1	51	0		244	1966	1976	49	1	48	0
70	1950	1955	60	1	69	0		245	1964	1972	60	0	44	1
71	1961	1967	62	1	57	0		246	1948	1956	48	1	69	0
72	1958	1968	54	1	56	0		247	1959	1964	60	1	59	0
73	1923	1932	50	1	66	1		248	1957	1961	63	1	62	0
74	1943	1950	75	1	70	1		249	1967	1970	57	1	48	0
75	1934	1940	70	1	52	1		250	1925	1932	82	1	60	1
76	1949	1956	70	1	68	0		251	1923	1926	82	1	67	1
77	1970	1979	53	1	45	0		252	1970	1978	28	1	46	0
78	1966	1974	55	1	50	0		253	1939	1946	84	1	60	1
79	1939	1948	85	0	69	1		254	1944	1952	47	1	69	1
80	1968	1976	52	1	48	0		255	1929	1932	82	1	85	1
81	1930	1937	68	1	85	1		256	1925	1939	62	1	85	1
82	1940	1948	67	1	76	0		257	1973	1974	50	1	50	0
83	1964	1970	60	1	54	0		258	1927	1940	75	1	68	1
84	1939	1945	57	1	68	1		259	1939	1947	82	1	60	1
85	1971	1979	53	0	45	1		260	1920	1926	79	0	39	1
86	1965	1971	58	1	53	0		261	1927	1932	67	1	60	1
87	1948	1953	76	0	50	1		262	1930	1936	76	1	37	1
88	1939	1944	65	1	80	1		263	1920	1928	87	0	41	1
89	1951	1958	73	0	55	1		264	1925	1933	75	1	37	1
90	1940	1947	68	1	75	1		265	1925	1935	87	1	36	1
91	1966	1976	56	1	48	1		266	1931	1937	86	1	35	1
92	1943	1952	81	1	72	0		267	1933	1941	85	1	75	1
93	1961	1968	59	1	56	0		268	1921	1926	68	0	54	1
94	1974	1979	50	0	41	1		269	1934	1942	77	1	60	1

Sr. No.	BYH	BYW	X1	SX1	X2	SX2		Sr. No.	BYH	BYW	X1	SX1	X2	SX2
95	1970	1976	54	1	40	1		270	1935	1944	36	1	38	1
96	1954	1961	61	1	63	0		271	1928	1937	79	0	46	1
97	1921	1929	89	1	70	1		272	1943	1948	65	0	35	1
98	1934	1940	69	1	73	1		273	1936	1941	45	1	70	1
99	1959	1964	65	0	59	1		274	1926	1931	76	1	54	0
100	1971	1981	51	1	43	0		275	1938	1944	79	1	69	1
101	1936	1940	70	1	80	1		276	1922	1929	71	1	60	1
102	1946	1954	77	1	70	0		277	1932	1938	90	0	53	0
103	1945	1950	75	1	71	1		278	1938	1944	63	0	48	1
104	1956	1965	68	0	55	1		279	1930	1940	77	0	52	1
105	1933	1939	73	1	69	1		280	1944	1953	76	1	39	1
106	1949	1956	72	1	68	0		281	1937	1942	56	1	60	1
107	1969	1977	52	1	47	0		282	1948	1952	66	1	67	1
108	1935	1944	75	1	70	1		283	1932	1941	78	1	75	1
109	1957	1964	53	1	60	0		284	1950	1956	43	1	40	1
110	1963	1968	61	0	52	1		285	1922	1932	85	1	64	0
111	1965	1972	55	1	52	0		286	1949	1956	65	1	66	1
112	1967	1976	56	1	48	0		287	1939	1946	70	1	56	1
113	1944	1949	70	1	75	0		288	1926	1933	38	1	65	1
114	1936	1944	77	1	70	1		289	1923	1928	57	0	70	0
115	1969	1975	45	1	49	0		290	1928	1933	80	1	72	1
116	1975	1982	49	0	32	1		291	1920	1929	76	0	69	0
117	1963	1968	60	1	56	0		292	1941	1951	69	1	47	1
118	1934	1939	81	1	72	1		293	1926	1933	84	1	65	1
119	1946	1951	60	1	70	1		294	1935	1940	85	1	58	1
120	1969	1977	55	0	44	1		295	1937	1947	47	1	52	1
121	1937	1947	84	1	77	0		296	1928	1936	70	1	85	1
122	1953	1960	58	1	64	0		297	1929	1939	78	0	60	1
123	1957	1964	67	0	59	1		298	1932	1939	53	1	61	0
124	1967	1975	56	1	49	0		299	1943	1952	74	1	48	1
125	1955	1965	69	0	55	1		300	1936	1943	73	1	58	1
126	1952	1961	66	1	59	1		301	1956	1963	54	1	38	1
127	1953	1962	61	1	62	0		302	1920	1929	80	1	73	1
128	1942	1947	64	1	70	1		303	1944	1950	80	0	72	1
129	1971	1979	53	1	45	0		304	1944	1950	69	1	52	1
130	1953	1959	64	1	51	1		305	1926	1933	62	1	70	1
131	1935	1943	70	1	62	1		306	1927	1936	81	1	67	0
132	1965	1973	59	1	51	0		307	1934	1943	84	1	60	1
133	1931	1936	79	1	88	0		308	1932	1941	90	1	62	1
134	1955	1963	69	0	57	1		309	1940	1947	55	1	64	1
135	1951	1956	57	1	68	0		310	1946	1955	60	1	49	1
136	1950	1957	74	0	65	1		311	1928	1934	38	1	71	1
137	1959	1969	56	1	43	1		312	1953	1963	46	1	43	1
138	1964	1969	60	0	55	1		313	1938	1948	75	1	59	0
139	1941	1950	81	1	74	0		314	1947	1957	67	0	50	0
140	1949	1959	75	0	63	1		315	1941	1944	77	1	75	1
141	1973	1979	51	0	44	1		316	1944	1950	36	1	60	1
142	1929	1936	75	0	73	1		317	1929	1939	74	1	71	1
143	1955	1963	60	1	61	1		318	1954	1962	68	0	48	1
144	1938	1946	72	1	60	1		319	1937	1943	66	1	68	1
145	1964	1970	55	1	54	1		320	1921	1930	86	0	81	1
146	1954	1962	57	1	62	0		321	1949	1959	74	1	52	1
147	1950	1958	63	1	65	1		322	1937	1947	87	0	64	1
148	1946	1955	70	1	69	0		323	1925	1935	56	0	78	1
149	1935	1940	80	1	60	1		324	1961	1967	60	1	57	0
150	1974	1979	50	0	40	1		325	1945	1952	62	1	61	0

Sr. No.	BYH	BYW	X1	SX1	X2	SX2		Sr. No.	BYH	BYW	X1	SX1	X2	SX2
151	1969	1975	49	1	49	0		326	1931	1939	80	1	75	1
152	1959	1965	63	1	59	0		327	1964	1971	53	1	42	1
153	1945	1949	79	1	75	0		328	1941	1950	81	1	73	1
154	1957	1965	66	1	59	0		329	1920	1925	72	1	89	0
155	1962	1969	62	1	51	1		330	1923	1931	41	1	84	1
156	1953	1958	71	0	60	1		331	1924	1929	68	1	86	0
157	1949	1958	65	1	53	1		332	1948	1953	51	0	62	1
158	1965	1973	59	0	49	1		333	1943	1949	59	1	67	0
159	1956	1965	63	1	57	1		334	1951	1960	53	0	56	1
160	1975	1985	49	0	32	1		335	1957	1965	67	0	58	1
161	1963	1972	58	1	52	0		336	1926	1931	47	1	86	1
162	1968	1974	48	1	50	0		337	1962	1971	58	1	46	1
163	1950	1956	62	1	68	1		338	1940	1949	65	1	65	1
164	1946	1953	78	0	68	1		339	1921	1929	53	1	91	0
165	1935	1944	79	1	76	1		340	1931	1936	70	1	85	1
166	1955	1960	69	0	57	1		341	1938	1944	79	1	76	1
167	1960	1964	44	1	60	0		342	1958	1964	51	1	57	1
168	1962	1969	56	1	55	0		343	1944	1951	40	1	71	1
169	1974	1978	50	0	44	1		344	1968	1978	56	0	44	1
170	1942	1950	60	1	74	0		345	1958	1963	30	1	60	1
171	1932	1941	89	1	83	0		346	1943	1951	80	0	72	0
172	1970	1976	54	1	48	0		347	1922	1931	51	1	93	0
173	1952	1960	52	1	64	0		348	1953	1962	38	1	62	0
174	1970	1975	54	0	46	1		349	1930	1939	65	1	85	1
175	1934	1939	61	1	55	1		350	1955	1962	59	1	62	1

2.4 Data Preprocessing

To prepare the data for analysis, several preprocessing steps are essential:

1. **Data Cleaning:** In this initial step, we identify and rectify any inconsistencies in the dataset, such as missing or incorrectly recorded birth years, ages, or survival statuses.

2. **Calculating Survival Time:** For each spouse, the survival time is calculated by determining the difference between the current age (or age at death) and the respective birth year. This calculated survival time is crucial for conducting traditional survival analysis methods, as it directly impacts the estimation of survival probabilities.

3. **Handling Censored Data:** The censored variable for both husbands and wives is defined such that it is set to 1 if the individual is deceased and 0 if they are still alive at the end of the study period. These censored observations are pivotal in survival analysis models, as they allow for the proper estimation of survival functions while accounting for incomplete data.

4. **Deriving New Variables:** After the initial data collection and cleaning, several new variables are derived to enhance the analysis:

    1. **X1:** Observed Life of Husband (Completed/Current)
    2. **X2:** Observed Life of Wife (Completed/Current)
    3. **X3: Remaining Age of Husband after Death of Wife:** This is calculated by subtracting wife's year of death from husband's year of death or current year for alive husbands.

4. **X4: Remaining Age of Wife after Death of Husband:** This is calculated by subtracting husband's year of death from wife's year of death or current year for alive wives.
5. **X5:** Age of husband who died before wife
6. **X6:** Age of wife who died before husband

- **Summary statistics based on raw data collected**

Variable	Status	N	Mean	SD	Min	Q1	Median	Q3	Max
X1	0	79	67.71	11.48	49	57	67	77	93
	1	271	63.6	13.06	28	54	63	75	90
X2	0	127	62.23	12.65	41	52	60	71	93
	1	223	59.91	12.64	31	51	60	70	86
X3	0	71	7.27	8.46	0	2	4	9	38
	1	66	11.73	10.37	0	4	9.5	17	46
X4	0	121	9.75	0.93	0	2	6	14	51
	1	95	13.69	11.54	0	4	12	19	51
X5	1	219	0.84	12.374	28	53	60	70	89
X6	1	131	54.71	10.808	31	47	55	63	79

0: Alive, 1: Died

Chapter 3: Methodology

This chapter outlines the methods used to analyse the survival data of spouses. It covers essential survival analysis concepts like the survival function, hazard function, and censoring types. The Kaplan-Meier estimator is used to estimate survival probabilities, while the Log-Rank Test compares survival curves between different groups. The Cox Proportional Hazards model helps identify significant covariates affecting survival times. Copula models, including Clayton, Gumbel, and Frank, are used to model the dependence between spouses' survival times. Sklar's theorem underpins the copula model approach. Parametric survival analysis and copula-based techniques are both applied for detailed analysis. The Log-Rank Test is particularly useful for comparing the survival distributions of two groups. These methods provide a comprehensive framework for studying spousal survival times and their dependencies. The following sections present the mathematical formulations and applications of these methods.

3.1 Basic Notations:

i) T : lifetime random variable
ii) $f(t)$: Death density function
iii) $\mathrm{F(t)}$: Cumulative event probability function
iv) $\mathrm{S(t)}$: Survival probability Function
v) $\mathrm{h(t)}$: Hazard function
vi) $\Delta\mathrm{t}$: A small interval of time
vii) λ : Hazard rate
viii) e_x : Expected life time

Type	Methods	Advantages	Disadvantages
Parametric	Exponential model Weibull model Gompertz model Makeham model	Easy to interpret, more efficient and accurate when the survival times follow a particular distribution.	When the distribution assumption is violated, it may be inconsistent and can give sub-optimal results.
Semi-parametric	Proportional Hazard model Cox model	The knowledge of the underlying distribution of survival times is not required.	The distribution of the outcome is unknown; not easy to interpret.
Non-parametric	Kaplan-Meier Nelson-Aalen	More efficient when no suitable theoretical distributions known.	Difficult to interpret; yields inaccurate estimates.

3.2 Censoring:

Censoring is a crucial concept in survival analysis because it accounts for incomplete data or situations where the event of interest (e.g., death, failure, recovery) hasn't been fully observed for all individuals in the study. In real-world survival analysis, not all subjects will experience the event during the study period, and therefore censoring is necessary to handle such cases. There are three main types of censoring: right censoring, left censoring, and interval censoring. Here's a detailed explanation of each type:

3.2.1 Right Censoring

Definition: Right censoring occurs when the event of interest has not yet occurred for some individuals by the end of the study or observation period. This is the most common form of censoring in survival analysis.

Implications for Survival Analysis: Right-censored data indicates that while the event has not occurred yet, it may occur after the study period. Traditional survival methods like the Kaplan-Meier estimator and the Cox Proportional Hazards Model are well-equipped to handle right-censored data. In these methods, censored individuals contribute to the calculation of the survival function up to the time when they are censored, but not after.

3.2.2 Left Censoring

Definition: Left censoring occurs when the event of interest has already occurred before the study begins, but the exact time of occurrence is unknown. This means that the individual experienced the event before entering the study, but the specific time of the event remains uncertain.

Implications for Survival Analysis: Left-censored data is less common than right-censored data. Some survival analysis methods, including modifications to the Cox Proportional Hazards model, can accommodate left-censored data, though it requires different adjustments compared to right censoring. For many practical applications, left censoring can be problematic as it introduces uncertainty about the early time periods of the study

3.2.3 Interval Censoring:

Definition: Interval censoring occurs when the event of interest is known to have occurred within a specific time interval, but the exact time is unknown. This typically arises when the event is only observed at discrete time points, such as during scheduled check-ups or follow-up visits.

Implication of Survival Analysis:

Interval censoring is more complex to handle than right or left censoring. Specific survival analysis methods, such as Turnbull's estimator, or parametric models, are often used to address interval-censored data. Interval-censored data requires exact event time is only known to lie within a certain range.

3.3 Survival Function Definition:

The survival function S(t) represents the probability that a subject survives beyond time t:

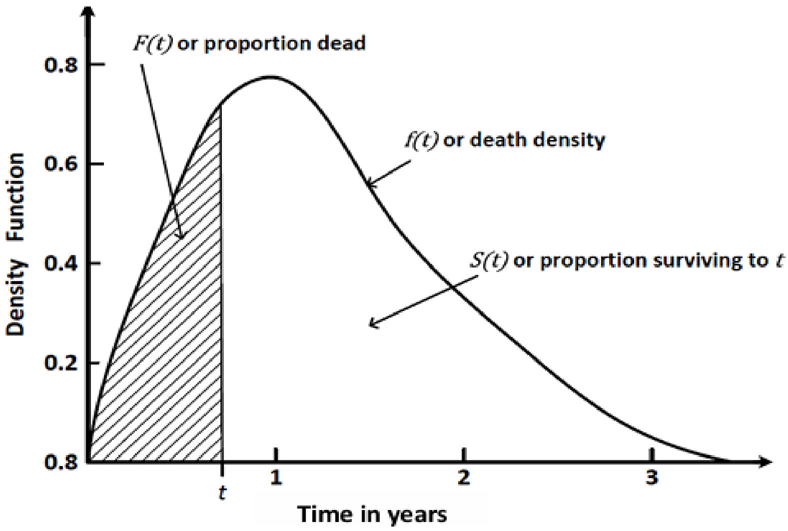
$$S(t) = P(T > t) = 1 - F(t)$$

On the other hand, the cumulative death distribution function F(t), which represents the probability that the event of interest occurs earlier than t.

$$F(t) = P[T \leq t]$$

For continuous cases F(t) can be obtained as $\mathrm{F(t)} = \frac{\mathrm{d}}{\mathrm{dx}}\mathrm{F(t)}$, for discrete case $\mathrm{f(t)} = \frac{[\mathrm{F(t+\Delta t)-F(t)}]}{\Delta \mathrm{t}}$.

This function can be plotted as



Hazard function is the instantaneous death rate or conditional failure rate. Also known as force of mortality in Actuarial Statistics. It does not indicate the chance or probability of the event of interest, but instead it is the rate of event at time t given that no event occurred before time t. Mathematical expression

given as:

$$h(t) = \lim_{\Delta t \to 0} P(t \leq T < t + \Delta t | T \geq t) / \Delta t$$

$$= \lim_{\Delta t \to 0} F(t + \Delta t) - F(t) / \Delta t . S(t)$$

$$= \frac{f(t)}{S(t)}$$

Like S(t), h(t) is non negative function. Survival function decrease over time, hazard function can have variety of shapes.

Where T is a non-negative random variable representing the time until the event of interest (e.g., death).

3.4 Kaplan-Meier Estimator

The Kaplan-Meier estimator is a non-parametric statistic used to estimate the survival function from lifetime data, particularly useful when dealing with censored data.

Kaplan-Meier Estimator Formula:

The Kaplan-Meier estimator $S\hat{}(t)$ is defined as:

$$\hat{S}(t) = \prod_{t_i \leq t} \left(1 - \frac{d_i}{n_i}\right)$$

where:

t_i are the ordered event times,

d_i is the number of events (deaths) at time t_i

c_i is the number of censored observations at time t_i,

n_i is the number of individuals at risk just before time t_i

Application to Spousal Data:

In our project, we will use the Kaplan-Meier estimator to:

1. Estimate the survival functions for husbands and wives separately
2. Compare the survival probabilities over time between husbands and wives.
3. Assess the impact of the death of one spouse on the survival probability of the other.

Calculation:

The Kaplan-Meier estimator is a step function that changes at each event time. The survival probability at time t is computed as:

Output:

Survival Curve: A step function that shows the estimated probability of survival over time.

Plot: Commonly visualized as a step plot with time on the x-axis and survival probability on the y-axis.

3.5 Cox Proportional Hazards Model

The Cox Proportional Hazards Model is a semi-parametric model that assesses the effect of covariates on the hazard rate without specifying the baseline hazard function.

Cox Model Formula:

$$h(t|X) = h_{0(t)}\exp(\beta^T X)$$

where:

$h(t|X)$ is the hazard function given covariates X,

$h_0(t)$ is the baseline hazard function,

β is a vector of regression coefficients,

X is a vector of covariates.

Partial Likelihood Function:

The Cox model is estimated using the partial likelihood:

$$L(\beta) = \prod_{i=1}^{D} \frac{\exp(\beta^T X_i)}{\sum_{j \in R_i} \exp(\beta^T X_j)}$$

where:

D is the set of individuals who experienced the event,

R_i is the risk set at time t_i (individuals still at risk just before time t_i).

Log-Partial Likelihood for Estimation:

$$\ell(\beta) = \sum_{i=1}^{D} \left[\beta^T X_i - \log \left(\sum_{j \in R_i} \exp(\beta^T X_j) \right) \right]$$

Assumptions:

Proportional Hazard Assumption: The hazard ratios are constant over time.

Linearity: The log hazard is a linear function of the covariates.

Application to Spousal Data:

Covariates: Age, gender, and an indicator variable for the death of the spouse.

Objective: Determine the effect of a spouse's death on the hazard rate of the surviving spouse.

Testing Proportional Hazards: Use Schoenfeld residuals to assess the validity of the proportional hazard assumption.

3.6 Log-rank Test:

The log-rank test is a statistical method used to compare the survival distributions of two or more groups. It assesses whether there are significant differences in the survival curves between these groups. Here's a more detailed explanation:

Purpose: The log-rank test evaluates the null hypothesis that there is no difference in the survival experience of the groups being compared. It is often used in clinical trials and survival studies to determine whether the survival curves for different treatment groups or populations differ significantly.

How It Works:

Estimate Survival Curves: For each group, survival curves are estimated using the Kaplan-Meier method. This involves calculating the probability of survival at different time points.

Compare Observed and Expected Events: The test compares the number of observed events (e.g., deaths) in each group to the number of events expected if the survival curves were the same across all groups.

Calculate the Test Statistic: The log-rank test statistic is computed based on the differences between observed and expected events. The test statistic follows a chi-squared distribution under the null hypothesis of no difference between survival curves.

Determine Significance: The test statistic is used to calculate a p-value, which indicates whether the observed differences in survival curves are statistically significant.

Null Hypothesis (H_0): There is **no difference** in the survival distributions between the groups being compared.

In mathematical terms: $H_0: S_{1(t)} = S_{2(t)} \ \forall$ t
Where:
$S_{1(t)}$ *and* $S_{2(t)}$ are the two survival functions of the two groups
Alternative Hypothesis (H_a): There is a **significant difference** in the survival distributions between the groups.
In mathematical terms: H_a: $H_a: S_{1(t)} \neq S_{2(t)}$ for at least some t.

- **Copula Model For to check dependency**

In survival analysis, understanding the dependency between two survival times is critical, particularly when modelling events such as the lifespans of spouses. Traditional survival models often struggle to capture complex dependencies, but copula models provide a robust framework to address this challenge. Copula models allow for the modeling of the **dependence structure** between the survival times of related individuals, such as married couples, offering insights into how the death of one partner may affect the survival of the other.

What is a Copula?

A **copula** is a mathematical function that links the joint distribution of two or more random variables to their marginal distributions. It allows for the separation of the marginal behaviour of each variable from their dependence structure. Formally, for two random variables T_1 andT_2 with marginal survival functions $S_{1(t_1)}$ and $S_{2(t_2)}$, a copula *C* combines these marginals into a joint survival function $S(t_1, t_2)$:

$$S(t_1, t_2) = C\left(S_{1(t_1)}, S_{2(t_2)};\ \theta\right)$$

where θ is the dependency parameter of the copula.

Types of Copulas: Several copulas exist to model different types of dependencies:

1. **Clayton Copula:** Models positive dependency, with a focus on the lower tail of the distribution. It is suitable for situations where events tend to occur together at low values of survival time.

2. **Gumbel Copula:** Captures dependency in the upper tail, making it suitable for cases where extreme values (long survival times) are of interest.

3. **Frank Copula:** Symmetric and can model both positive and negative dependencies. It is appropriate for a wide range of dependency structures.

Sklar's Theorem: Sklar's theorem forms the theoretical foundation of copula models. It states that for any multivariate distribution $F(t_1,t_2)$ with continuous marginals $F_1(t_1)$ and $F_2(t_2)$, there exists a unique copula *C* such that:

$$F(t_1, t_2) = C\left(F_{1(t_1)}, F_{2(t_2)}\right).$$

This decomposition allows the dependency structure and marginal behaviour to be model independently.

Advantages of Copula Models:

Flexibility: Copulas can model a wide range of dependency structures.
Tail Dependence: Some copulas, such as Clayton and Gumbel, are specifically designed to capture dependencies in the tails of the distribution.
Application to Survival Data: Copulas are well-suited for censored survival data, where marginal survival functions are derived separately.

Chapter 4: Data Analysis

This chapter focuses on analysing the survival data using a combination of statistical methods to uncover meaningful insights. The **Kaplan-Meier estimator** is used to estimate the survival probabilities and visualize the survival patterns over time. This non-parametric method provides a clear picture of how survival changes across different time periods. To compare survival distributions between groups, the **log-rank test and cox model** is applied, helping to identify significant differences in survival trends.

A key focus of the analysis is understanding the relationship between the survival times of husbands and wives. For this, **copula models** are employed, which are powerful tools for studying dependencies between two variables. These models allow us to explore how the death of one partner affects the survival probability of the other, providing valuable insights into the interdependence of their survival times. The results of the copula models, combined with Kaplan-Meier curves and statistical tests, offer a comprehensive understanding of survival patterns and relationships. The visualizations and detailed interpretations presented in this chapter aim to clarify how survival times are influenced by various factors, including the relationship between partners. These findings contribute to a deeper understanding of survival dynamics and their implications.

4.1 Survival analysis on observed life of spouses

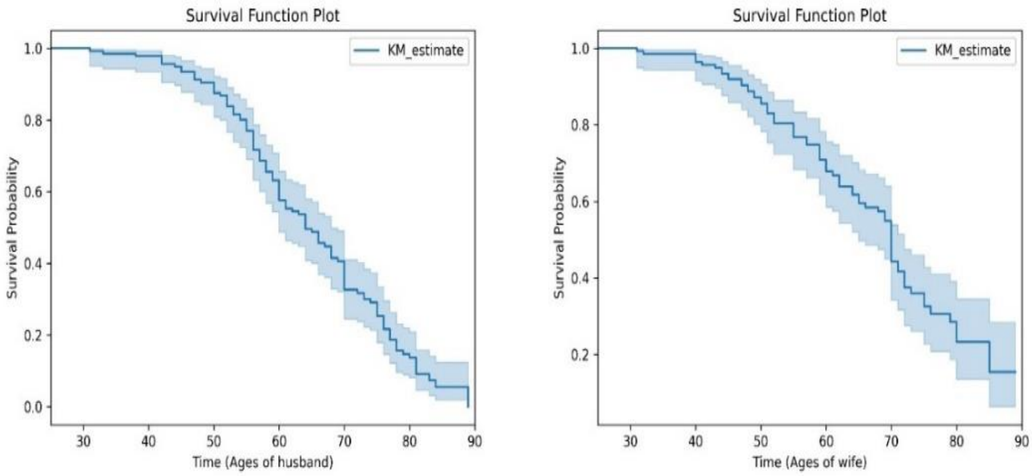


Fig.1 Kaplan-Meier curve of (X1) and X2)

The Kaplan-Meier plots show the survival probabilities for husbands (left) and wives (right) based on their ages. The survival curve for husbands declines faster, meaning their chances of survival decrease more quickly as they age. In comparison, the wives' survival curve declines more gradually, showing that they have a higher chance of survival at older ages. The shaded areas around the curves represent confidence intervals, which indicate the uncertainty in the survival estimates. Overall, the plots suggest that wives tend to live longer than husbands.

- **Joint Survival Curves using r programming**

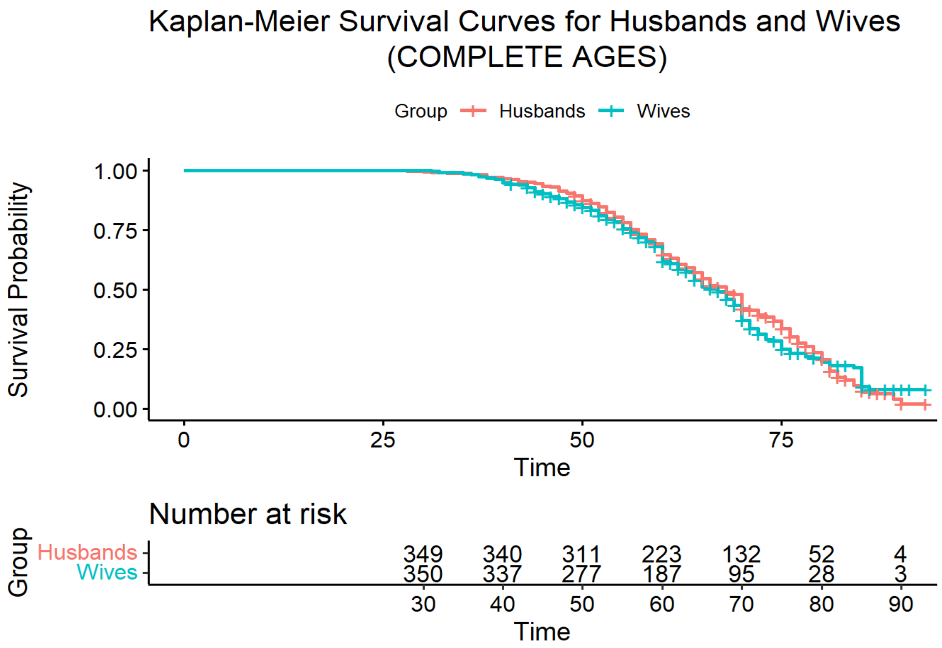


Fig.2 Joint Kaplan Curve of X1 and X2

The "Number at Risk" shows the percentage of husbands and wives still alive at different time points. At time 30, 100% of both groups are being tracked. By time 40, around 97% of husbands and 96% of wives remain. At time 50, 89% of husbands and 79% of wives are still alive. By time 60, survival drops more sharply, with 64% of husbands and 53% of wives still being tracked. By time 80, only 15% of husbands and 8% of wives remain. Finally, by time 90, just 1% of husbands and wives are still alive. This shows a steady decline in survival as time progresses, with most individuals not reaching age 90.

- **Log rank Test**

```
Call:
survdiff(formula = Surv(time, event) ~ group, data = data_pair1)

           N Observed Expected (O-E)^2/E (O-E)^2/V
group=X1 350      271      276     0.103     0.249
group=X2 350      223      218     0.131     0.249

 Chisq= 0.2  on 1 degrees of freedom, p= 0.6
```

In above test p value is 0.6 which is greater than 0.05(alpha) so we can conclude that there is no significance difference in above survival curve for complete ages between husband and wife.

The following tables gives the survival probabilities of husbands (Left) and wives (Right)

Call: survfit(formula = surv_obj1 ~ 1, data = data)

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
28	350	1	0.9971	0.00285	0.99157	1.0000
30	349	1	0.9943	0.00403	0.98642	1.0000
31	348	1	0.9914	0.00493	0.98182	1.0000
33	347	1	0.9886	0.00568	0.97750	0.9998
36	346	2	0.9829	0.00694	0.96935	0.9966
38	344	4	0.9714	0.00891	0.95413	0.9890
40	340	2	0.9657	0.00973	0.94684	0.9850
41	338	1	0.9629	0.01011	0.94325	0.9829
42	337	3	0.9543	0.01116	0.93265	0.9764
43	334	1	0.9514	0.01149	0.92917	0.9742
44	333	2	0.9457	0.01211	0.92227	0.9698
45	331	4	0.9343	0.01324	0.90868	0.9606
46	327	1	0.9314	0.01351	0.90532	0.9583
47	326	6	0.9143	0.01496	0.88542	0.9441
48	320	3	0.9057	0.01562	0.87561	0.9369
49	317	4	0.8943	0.01644	0.86265	0.9271
50	311	7	0.8742	0.01774	0.84007	0.9096
51	301	4	0.8625	0.01843	0.82717	0.8994
52	294	5	0.8479	0.01925	0.81097	0.8864
53	289	8	0.8244	0.02043	0.78532	0.8654
54	278	7	0.8036	0.02137	0.76284	0.8466
55	269	8	0.7797	0.02234	0.73717	0.8248
56	260	9	0.7528	0.02331	0.70843	0.7998
57	248	7	0.7315	0.02399	0.68596	0.7801
58	238	7	0.7100	0.02463	0.66333	0.7599
59	231	6	0.6915	0.02511	0.64404	0.7426
60	223	15	0.6450	0.02614	0.59579	0.6983
61	204	4	0.6324	0.02638	0.58274	0.6863
62	199	8	0.6070	0.02681	0.55663	0.6618
63	191	5	0.5911	0.02703	0.54040	0.6465
64	184	6	0.5718	0.02727	0.52077	0.6278
65	177	8	0.5460	0.02753	0.49459	0.6027
66	166	9	0.5164	0.02774	0.46474	0.5737
67	155	2	0.5097	0.02778	0.45804	0.5672
68	148	6	0.4890	0.02791	0.43728	0.5469
69	139	3	0.4785	0.02796	0.42669	0.5365
70	132	16	0.4205	0.02808	0.36889	0.4793
71	115	2	0.4132	0.02807	0.36166	0.4720
72	111	5	0.3946	0.02801	0.34330	0.4535
73	105	3	0.3833	0.02795	0.33223	0.4422
74	100	4	0.3679	0.02787	0.31719	0.4268
75	95	8	0.3370	0.02759	0.28700	0.3956
76	85	9	0.3013	0.02711	0.25257	0.3594
77	73	6	0.2765	0.02670	0.22884	0.3341
78	66	4	0.2598	0.02637	0.21290	0.3169
79	60	6	0.2338	0.02577	0.18836	0.2902
80	52	6	0.2068	0.02504	0.16312	0.2622
81	42	10	0.1576	0.02343	0.11774	0.2109
82	30	5	0.1313	0.02227	0.09417	0.1831
83	24	2	0.1204	0.02172	0.08451	0.1714
84	21	4	0.0974	0.02038	0.06467	0.1468
85	17	4	0.0745	0.01853	0.04576	0.1213
86	12	1	0.0683	0.01800	0.04075	0.1145
87	10	1	0.0615	0.01745	0.03525	0.1072
89	6	2	0.0410	0.01659	0.01854	0.0906
90	4	2	0.0205	0.01318	0.00581	0.0723

Call: survfit(formula = surv_obj2 ~ 1, data = data)

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
31	350	1	0.9971	0.00285	0.9916	1.000
32	349	2	0.9914	0.00493	0.9818	1.000
35	347	2	0.9857	0.00634	0.9734	0.998
36	345	1	0.9829	0.00694	0.9694	0.997
37	344	3	0.9743	0.00846	0.9578	0.991
38	341	2	0.9686	0.00933	0.9505	0.987
39	339	2	0.9629	0.01011	0.9432	0.983
40	337	5	0.9486	0.01181	0.9257	0.972
41	332	2	0.9429	0.01241	0.9189	0.967
42	329	1	0.9400	0.01270	0.9154	0.965
43	328	4	0.9285	0.01377	0.9019	0.956
44	323	6	0.9113	0.01521	0.8819	0.942
45	315	3	0.9026	0.01587	0.8720	0.934
46	309	4	0.8909	0.01671	0.8588	0.924
47	302	3	0.8821	0.01730	0.8488	0.917
48	296	5	0.8672	0.01825	0.8321	0.904
49	285	4	0.8550	0.01898	0.8186	0.893
50	277	3	0.8457	0.01951	0.8083	0.885
51	269	4	0.8332	0.02021	0.7945	0.874
52	262	7	0.8109	0.02135	0.7701	0.854
53	252	5	0.7948	0.02211	0.7526	0.839
54	241	3	0.7849	0.02256	0.7419	0.830
55	235	9	0.7549	0.02382	0.7096	0.803
56	223	4	0.7413	0.02433	0.6951	0.791
57	214	7	0.7171	0.02520	0.6693	0.768
58	204	5	0.6995	0.02578	0.6507	0.752
59	197	5	0.6817	0.02632	0.6321	0.735
60	187	18	0.6161	0.02797	0.5637	0.673
61	166	2	0.6087	0.02812	0.5560	0.666
62	162	6	0.5861	0.02854	0.5328	0.645
63	151	3	0.5745	0.02876	0.5208	0.634
64	147	9	0.5393	0.02929	0.4849	0.600
65	131	7	0.5105	0.02968	0.4555	0.572
66	124	2	0.5023	0.02977	0.4472	0.564
67	121	3	0.4898	0.02988	0.4346	0.552
68	115	7	0.4600	0.03012	0.4046	0.523
69	105	6	0.4337	0.03025	0.3783	0.497
70	95	14	0.3698	0.03023	0.3151	0.434
71	78	7	0.3366	0.03001	0.2827	0.401
72	68	5	0.3119	0.02977	0.2586	0.376
73	58	4	0.2904	0.02960	0.2378	0.355
74	54	1	0.2850	0.02953	0.2326	0.349
75	49	6	0.2501	0.02915	0.1990	0.314
76	40	3	0.2313	0.02891	0.1811	0.296
78	34	2	0.2177	0.02876	0.1681	0.282
79	32	1	0.2109	0.02866	0.1616	0.275
80	28	2	0.1959	0.02852	0.1472	0.261
81	26	2	0.1808	0.02825	0.1331	0.246
84	20	1	0.1717	0.02824	0.1244	0.237
85	19	9	0.0904	0.02466	0.0530	0.154
86	8	1	0.0791	0.02403	0.0436	0.143

This data represents the Kaplan-Meier survival probabilities over time for two groups: **Husband** and **Wife**.

Timeline: The timeline represents the time points (years) when survival events (death) were observed. At time 0, the survival probability is always 1.0 (100%) since no events have occurred yet.

1. **Husband Group**:
   - The survival probability starts at 1.0 and decreases as events occur over time.
   - At **timeline 31**, the survival probability drops to 0.9927, meaning that approximately 99.27% of husbands have "survived" (not experienced the event of interest).

- The survival probability continues to decline, with a sharp drop observed between **timeline 70** (32.67%) and **timeline 80** (13.75%), indicating a high frequency of events during this period.
- At the final timeline point (89), the survival probability reaches **0.0000**, indicating that all members of this group have experienced the event by this time.

2. **Wife Group**:
   - Similar to the husbands, the survival probability starts at 1.0 and decreases over time.
   - At **timeline 31**, the survival probability is 0.9927, the same as for the husbands, but it diverges slightly after that point.
   - The decline in survival probability is steadier compared to the husbands, with fewer sharp drops.
   - At **timeline 80**, the survival probability is 23.27%, which is higher than the husbands' probability at the same time point.
   - The final survival probability at **timeline 89** is 0.1552, indicating that some members of the wife group still "survive" (haven't experienced the event).

4.2 Survival analysis of remaining life of spouse after the death of other spouse

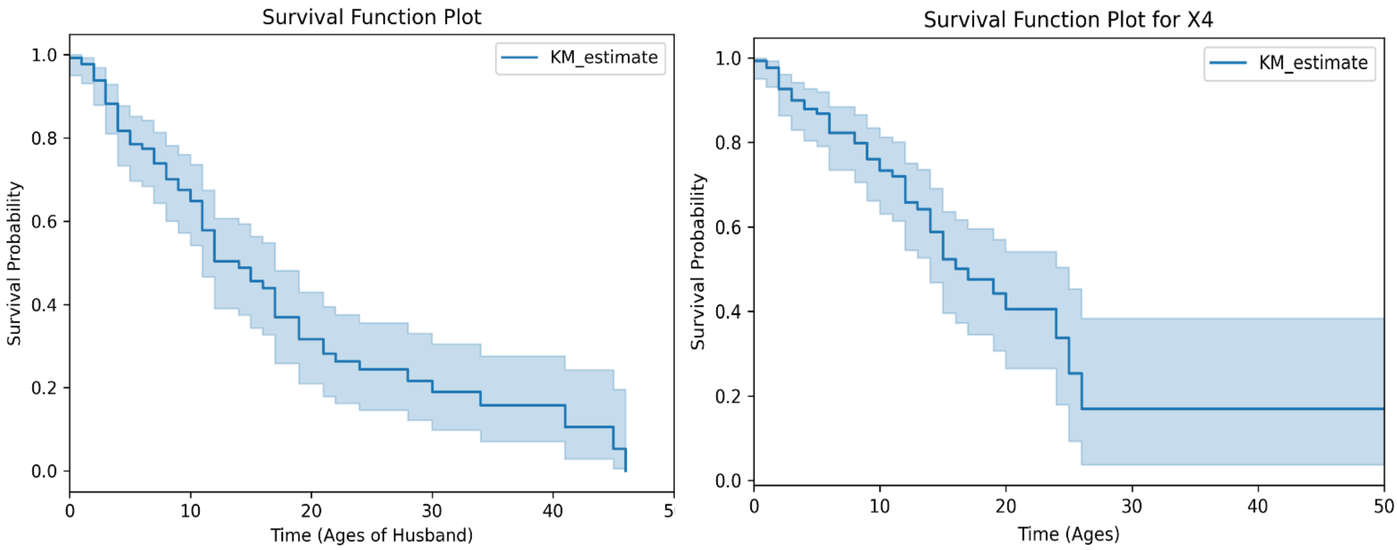


Fig.3 Survival Function Plot of X3 and X4

We can observe that in the above survival function plot of Remaining Age of Husband after Death of Wife and Remaining Age of Wive after Death of Husband the survival probability of X4 is greater than X3 so we conclude that the life expectancy of wives is more than husband after death of spouse.

Joint Survival Curves X3 and X4:

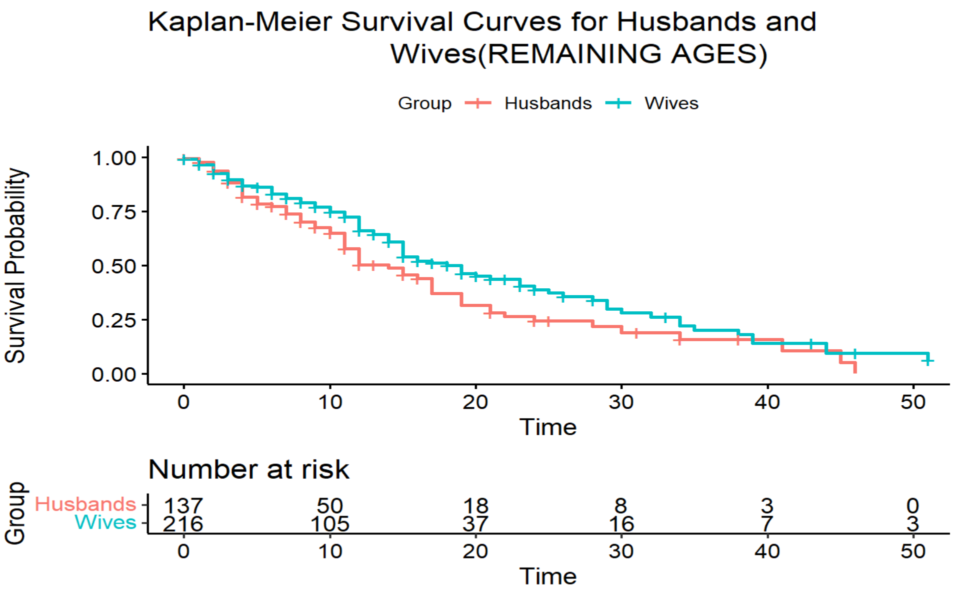


Fig.4 Joint Kaplan Curve of X3 and X4

This survival analysis shows how long husbands and wives live after the death of their spouse. At time 0, 100% of both husbands and wives are still alive. However, by time 10, only 36.5% of husbands and 48.6% of wives remain. By time 20, survival decreases further, with only 13.1% of husbands and 17.1% of wives alive. At time 30, the numbers drop to 5.8% of husbands and 7.4% of wives. By time 40, only 2.2% of husbands and 3.2% of wives are still alive. Finally, at time 50, no husbands survive (0%), while just 1.4% of wives remain. Overall, wives tend to live longer after their husbands' death compared to husbands after their wives' death.

Analysis of Remaining Age of Husband after Death of Wife and Remaining Age of Wive after Death of Husband

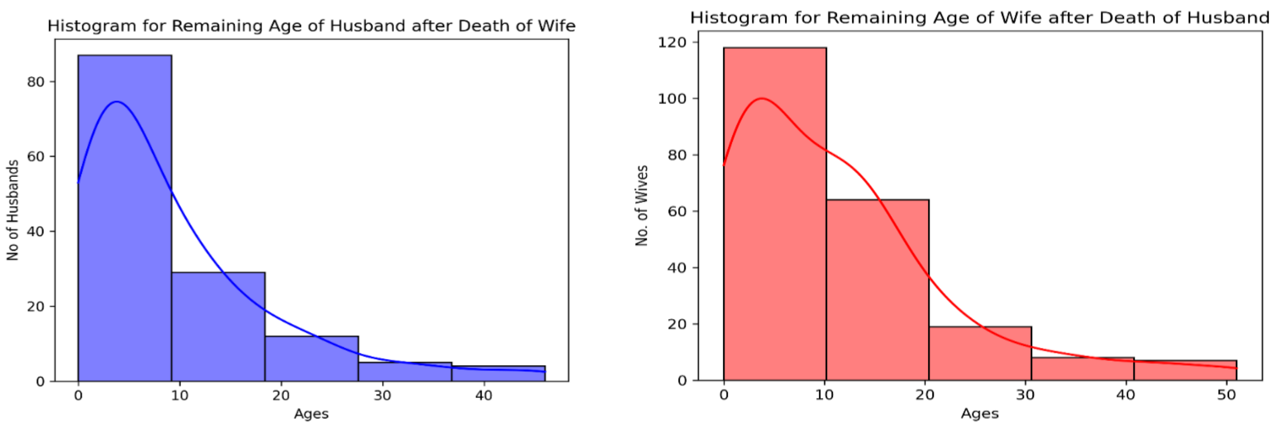


Fig.5 Histogram of X3 and X4

These histograms display the remaining ages of husbands and wives after the death of their spouses:

Left Histogram (Husbands):

- This shows the distribution of the remaining lifespan of husbands after the death of their wives.
- Most husbands live for a short duration after their wives' death, with the majority surviving less than 10 years.
- The frequency drops sharply after 10 years, with very few husbands living beyond 20 years.
- The distribution is right-skewed, indicating that the majority of husbands have a shorter remaining lifespan.

Right Histogram (Wives):

- This shows the distribution of the remaining lifespan of wives after the death of their husbands.
- Similar to husbands, most wives live less than 10 years after their husbands' death.
- A small proportion of wives live longer, with a few surviving more than 30 years.
- The distribution is also right-skewed but extends further than the husbands' distribution, indicating that wives may live longer on average after their husbands' death.

4.3 Log rank, Cox and Copula tests

It compares the equality of survival distribution, its non-parametric test.

Log rank test for variable X3 and X4

```
Call:
survdiff(formula = Surv(time, event) ~ group, data = data_pair2)

n=353, 347 observations deleted due to missingness.

           N Observed Expected (O-E)^2/E (O-E)^2/V
group=X3 137       66     54.3      2.53      4.02
group=X4 216       95    106.7      1.28      4.02

 Chisq= 4  on 1 degrees of freedom, p= 0.04
```

In above test p value is 0.04 which is less than 0.05(alpha) so we can conclude that there significance difference in above survival curve of Remaining Age of Husband after Death of Wife and Remaining Age of wife after Death of husband**.**

Cox Test for X3 and X4:

```
Call:
coxph(formula = Surv(time, event) ~ group, data = data_pair2)

  n= 353, number of events= 161
   (347 observations deleted due to missingness)

            coef exp(coef) se(coef)      z Pr(>|z|)
groupX4 -0.3214    0.7251   0.1611 -1.995   0.0461 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

        exp(coef) exp(-coef) lower .95 upper .95
groupX4    0.7251      1.379    0.5288    0.9944

Concordance= 0.539  (se = 0.023 )
Likelihood ratio test= 3.89  on 1 df,   p=0.05
Wald test            = 3.98  on 1 df,   p=0.05
Score (logrank) test = 4.01  on 1 df,   p=0.05
```

In above test p value is 0.05 which is equal to 0.05(alpha) so we can conclude that there is dependence between the Remaining Age of Husband after Death of Wife and Remaining Age of Wife after Death of husband.

Copula test

It used to test dependency between survival distribution.

```
Dependency Parameters for Each Copula:
Clayton: 0.9140394355427677
Gumbel: 1.457019717771384
Frank: 3.0739832822731943
```

```
**Dependency Parameters Interpretation:**
Clayton Copula (θ = 0.9140): Indicates moderate to strong positive dependence, especially in lower survival times. This suggests that if one spouse's survival probability decreases (e.g., nearing death), the other is also likely to experience a similar decrease, emphasizing joint early deaths.

Gumbel Copula (θ = 1.4570): Shows moderate to strong positive dependence, particularly in higher survival times. This highlights that spouses are likely to survive longer together, emphasizing joint longevity.

Frank Copula (θ = 3.0739): Represents moderate positive dependence across all survival times. It captures a balanced relationship without emphasizing either early joint deaths or joint long-term survival.
```

From above results on copula from the three type we conclude that there is dependence in survival functions Remaining Age of Husband after Death of Wife and Remaining Age of Wife after Death of husband.

The following tables gives the survival probabilities of Remaining Age of Husband after Death of Wife (Left) and Remaining Age of Wife after Death of husband (Right)

Call: survfit(formula = surv_obj3 ~ 1, data = data)

213 observations deleted due to missingness

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
0	137	1	0.9927	0.00727	0.97855	1.000
1	129	2	0.9773	0.01296	0.95224	1.000
2	122	5	0.9373	0.02150	0.89606	0.980
3	103	6	0.8827	0.02963	0.82646	0.943
4	94	7	0.8169	0.03637	0.74866	0.891
5	77	3	0.7851	0.03933	0.71169	0.866
6	70	1	0.7739	0.04033	0.69874	0.857
7	66	3	0.7387	0.04331	0.65852	0.829
8	59	3	0.7011	0.04622	0.61616	0.798
9	54	2	0.6752	0.04802	0.58733	0.776
10	50	2	0.6482	0.04975	0.55764	0.753
11	46	5	0.5777	0.05340	0.48200	0.692
12	39	5	0.5037	0.05589	0.40521	0.626
14	32	1	0.4879	0.05631	0.38913	0.612
15	31	2	0.4564	0.05691	0.35748	0.583
16	27	1	0.4395	0.05726	0.34049	0.567
17	25	4	0.3692	0.05789	0.27151	0.502
19	21	3	0.3165	0.05707	0.22223	0.451
21	18	2	0.2813	0.05589	0.19057	0.415
22	15	1	0.2625	0.05522	0.17385	0.396
24	14	1	0.2438	0.05436	0.15747	0.377
28	9	1	0.2167	0.05466	0.13218	0.355
30	8	1	0.1896	0.05412	0.10837	0.332
34	6	1	0.1580	0.05354	0.08134	0.307
41	3	1	0.1053	0.05589	0.03724	0.298
45	2	1	0.0527	0.04656	0.00931	0.298
46	1	1	0.0000	NaN	NA	NA

134 observations deleted due to missingness

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
0	216	2	0.9907	0.00652	0.9780	1.000
1	201	5	0.9661	0.01260	0.9417	0.991
2	186	8	0.9245	0.01876	0.8885	0.962
3	168	5	0.8970	0.02187	0.8552	0.941
4	154	5	0.8679	0.02474	0.8207	0.918
5	142	1	0.8618	0.02531	0.8136	0.913
6	134	5	0.8296	0.02815	0.7763	0.887
7	124	3	0.8096	0.02976	0.7533	0.870
8	120	3	0.7893	0.03123	0.7304	0.853
9	115	3	0.7687	0.03260	0.7074	0.835
10	105	3	0.7468	0.03404	0.6829	0.817
11	98	3	0.7239	0.03547	0.6576	0.797
12	90	8	0.6596	0.03893	0.5875	0.740
13	79	2	0.6429	0.03970	0.5696	0.726
14	74	4	0.6081	0.04118	0.5325	0.694
15	63	7	0.5405	0.04381	0.4611	0.634
16	55	2	0.5209	0.04437	0.4408	0.616
17	49	1	0.5103	0.04472	0.4297	0.606
18	43	1	0.4984	0.04523	0.4172	0.595
19	41	3	0.4619	0.04656	0.3791	0.563
20	37	1	0.4494	0.04695	0.3662	0.552
21	34	1	0.4362	0.04739	0.3526	0.540
23	28	2	0.4051	0.04886	0.3198	0.513
24	25	1	0.3889	0.04952	0.3030	0.499
25	23	1	0.3720	0.05017	0.2855	0.485
26	22	1	0.3550	0.05066	0.2684	0.470
28	20	1	0.3373	0.05114	0.2506	0.454
29	18	2	0.2998	0.05187	0.2136	0.421
30	16	1	0.2811	0.05190	0.1957	0.404
32	15	1	0.2623	0.05172	0.1783	0.386
34	13	2	0.2220	0.05103	0.1415	0.348
35	11	1	0.2018	0.05022	0.1239	0.329
38	10	1	0.1816	0.04909	0.1069	0.308
39	9	2	0.1413	0.04573	0.0749	0.266
44	6	2	0.0942	0.04085	0.0402	0.220
51	3	1	0.0628	0.03740	0.0195	0.202

This data represents the Kaplan-Meier survival probabilities over time for two groups Remaining Age of Husband after Death of Wife and Remaining Age of Wife after Death of husband**.** We can see that wives have quite more survival probability than husbands.

- **Survival curves for Age of husband who died before wife and Age of wife who died before**

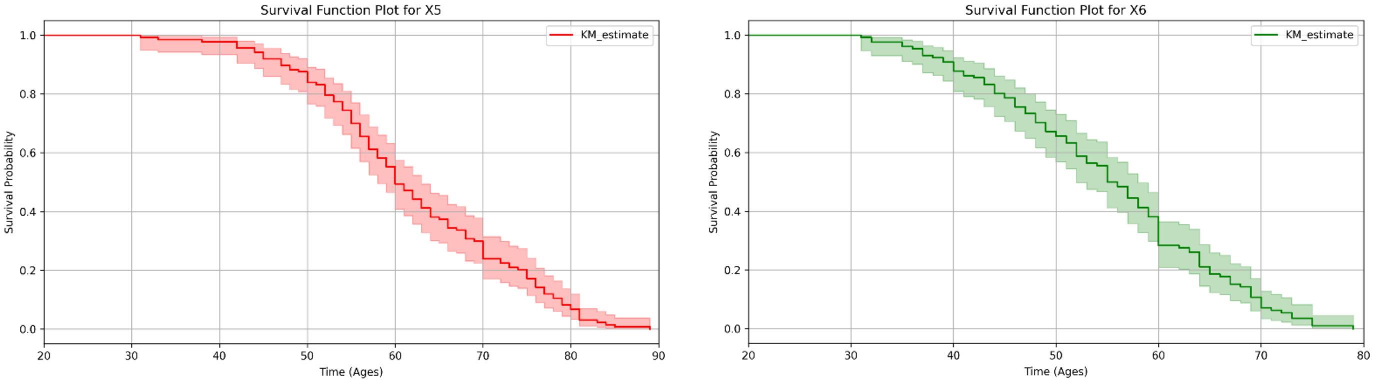


Fig.6 Survival Function Plot of X5 and X6

The plots show survival probabilities for husbands (X5) and wives (X6) who died before their spouses. Husbands tend to die at older ages, with a sharp decline in survival probability after age 60, and some living up to 90 years. Wives show a gradual decline in survival probability, with a sharper drop after age 70, and most passing away by 80 years. Overall, this suggests that wives generally live longer, but when they die first, their lifespan tends to end earlier than husbands who die first.

- **Joint Survival Curves for X5 and X6**

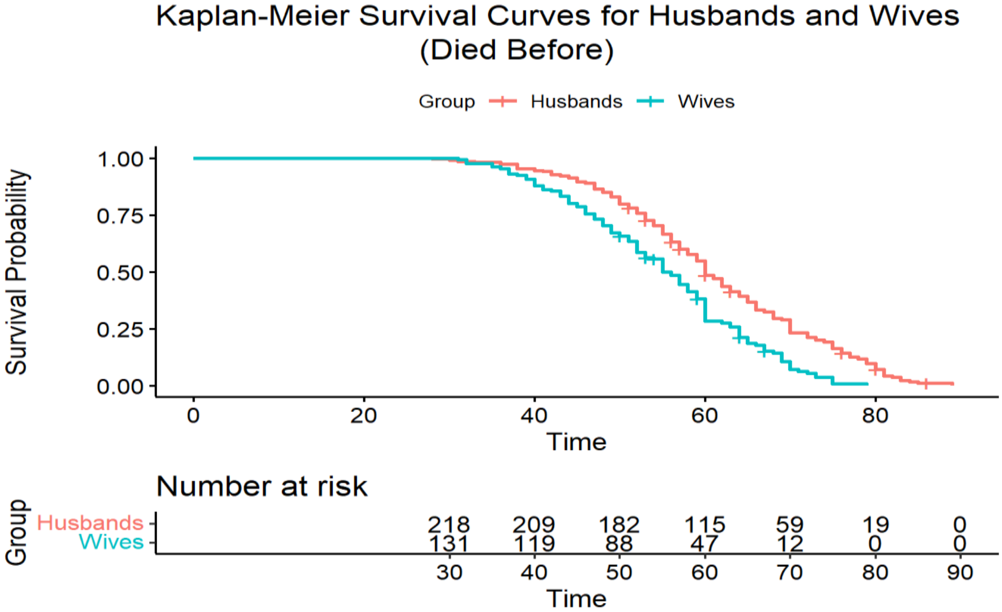


Fig.7 Joint Kaplan survival Curve of X5 and X6

The red curve represents the survival probabilities of husbands before their wives' death (SX2=1), while the blue curve represents wives' survival probabilities before their husbands' death (SX1=1).

Wives generally outlive husbands, as shown by the blue curve being above the green curve. Husbands' survival probabilities decline more steeply, indicating shorter lifespans after their wives' death. The shaded areas around the curves show confidence intervals, with some overlap, suggesting variability but a consistent trend. Both probabilities approach zero around age 90, reflecting low survival chances at advanced ages.

- **Joint Survival Curves for X1, X3 and X5:**

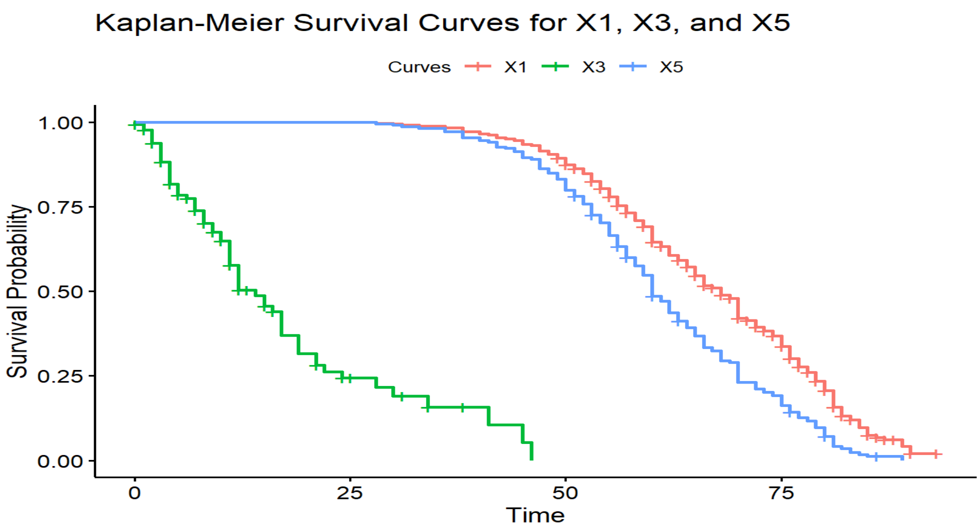


Fig.8 Joint Kaplan survival Curve of Complete Life of Husband, Remaining Age of Husband after Death of Wife, Age of husband who died before wife.

- **Joint Survival Curves for X2, X4 and X6:**

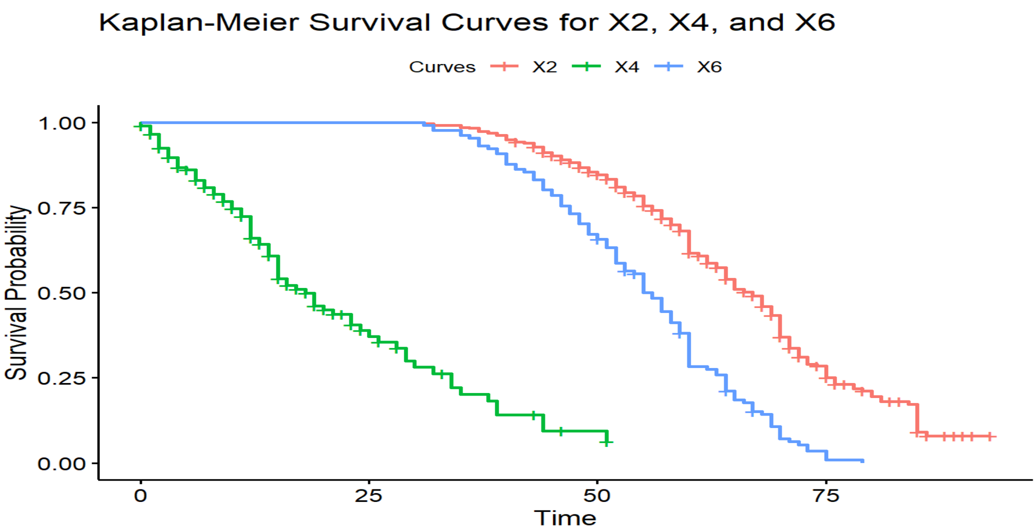


Fig.9 Joint Kaplan survival Curve of Complete Life of Wife, Remaining Age of Wive after Death of Husband, Age of wife who died before husband.

Subgroup Analysis

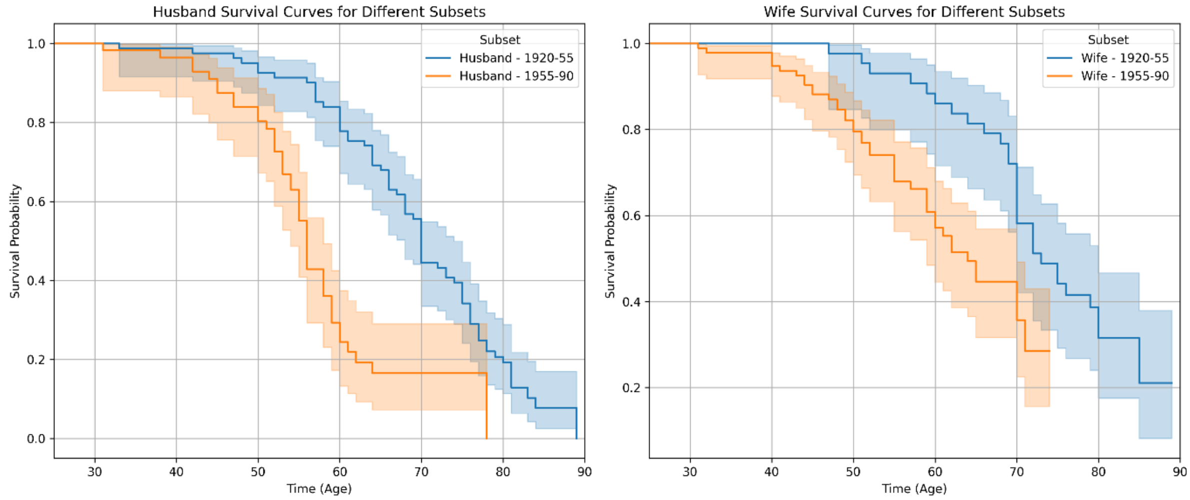


Fig.10 Joint Survival Function Plot of subgroup of husbands and wives

The Survival Function Plot of the subset of husbands and wives based on their birth years (1920–1955 and 1955–1990) shows that the survival probability of both husbands and wives is higher in the 1920–1955 subset compared to the 1955–1990 subset.

Log rank test between two groups of husbands:

```
<lifelines.StatisticalResult: logrank_test>
               t_0 = -1
 null_distribution = chi squared
degrees_of_freedom = 1
         test_name = logrank_test

---
 test_statistic      p  -log2(p)
          38.14 <0.005     30.50
```

In this study, we conducted a log-rank test to compare the survival distributions of husbands frow two subgroups based on their birth year :1920-55 and 1955-90. The p value very small, we reject the null hypothesis and conclude that there is significance in survival probability between the two subgroup of husbands his result suggests that survival patterns of husbands differ significantly depending on their birth cohort, possibly due to generational differences in lifestyle, healthcare access, or other socio-economic factors.

Log rank test between two groups of wives:

```
<lifelines.StatisticalResult: logrank_test>
               t_0 = -1
 null_distribution = chi squared
degrees_of_freedom = 1
         test_name = logrank_test

---
 test_statistic      p  -log2(p)
           9.91 <0.005      9.25
```

The result from the log-rank test compares the survival distributions of two groups (e.g., wives from the periods 1920-1955 and 1955-1990). Since the p-value is very small (less than 0.005), we reject the null hypothesis, which suggests that there is a statistically significant difference between the survival distributions of the two groups.

Descriptive Statistics of the age difference of husband and wife

count	mean	std	min	25%	50%	75%	max
350	7.088571	1.892720	1	6	7	8.75	14

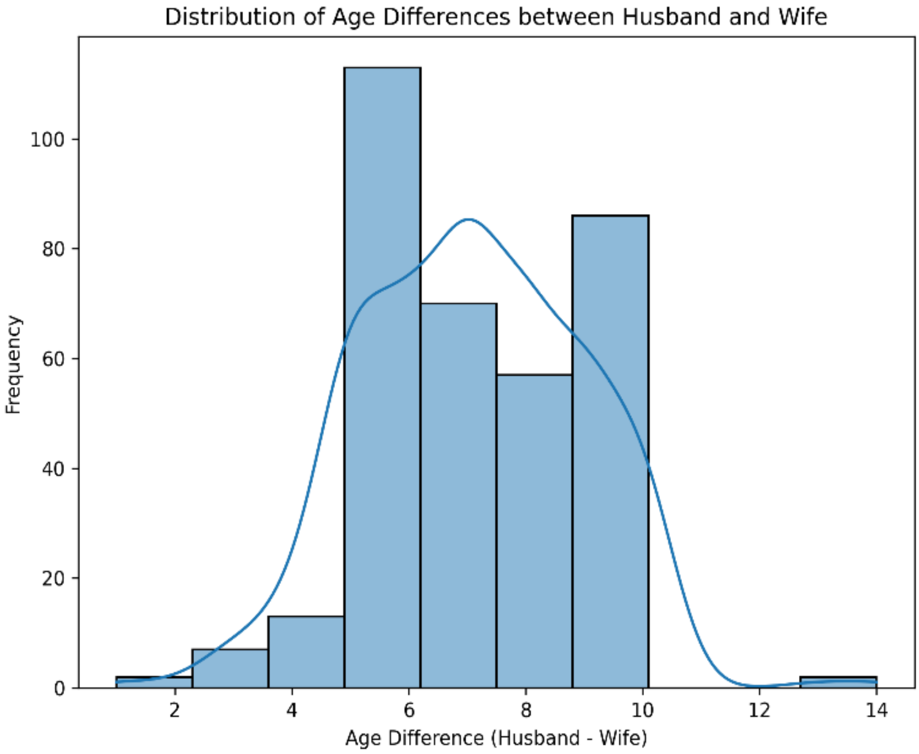


Fig.11 Histogram for age differences between husband and wife

We can see in this plot that the age difference between life is minimum 1 and maximum 14. The mean of age difference is 7.088 and standard deviation 1.89.

- **Summary table of variables of fitted model:**

Variables Name	No. obs.	Min	Median	Mean	Max	SD
X1 (observed life of husband)	350	28	61.5	65.97	90	17.07
X2 (observed life of wife)	350	31	58	64.59	86	15.41
X3 (rem. Life of husband After death of wife)	216	0	14	16.03	46	13.35
X4 (rem. Life of wife After death of husband)	137	0	17.5	20.46	51	13.05
X5 (husband died before wife)	219	28	59.5	60.04	89	16.36
X6 (wife died before husband)	131	31	53.5	53.7	79	13.04

Chapter 5: Review of Research Paper

5.1 Review of "The Linked Lives, Dependent Demise? Survival Analysis of Husbands and Wives."

The paper by Ken R. Smith and Cathleen D. Zick explores an interesting concept of whether couples share not only their lives but also the risks of mortality. The authors ask whether shared characteristics, like lifestyle, financial resources, or family, affect both spouses similarly in terms of death risks.

Abstract

The basis of this research is on the fact that a married person lives longer than his or her single or divorced counterparts. However, the question that this research takes deep into consideration is to know whether the characteristics which explain the mortality rate of the husbands and the wives are the same and different, respectively. What the authors use to estimate how diverse risk factors such as smoking, poverty, and children affect the couple is through the use of the paired hazard rate model.

Interestingly, they find that some factors, such as smoking, risk-avoidance behaviours, and poverty, affect both husbands and wives similarly. However, they also note that divorce impacts spouses differently, suggesting that while some risks are shared, others are experienced more individually.

Literature Review:

Marital Status and Mortality Relationship:

A variety of studies have shown that married people tend to live longer than the divorced, the widowed, or the never-married. Researchers usually one of three approaches to the study of this relationship:

Death Certificates and Population Data: Some studies used death certificates in combination with population data to calculate death rates for different marital statuses. In any case, these studies report uniformly that married people enjoy large advantages in longevity.

Longitudinal health data: In another, a baseline of demographic and health data are collected, after which follow-ups are made of those followed. Even in longitudinal data when unhealthy habits of smoking and drinking have controlled for, married individuals enjoy more survival.

Micro-Level Panel Data: The third approach employs detailed data on married couples to look at how marital status and changes in marital status (such as divorce) affect mortality risks. Such studies again indicate that marriage is related to longer life, particularly for men.

Explanations for Mortality Differences: Researchers offer two major explanations for why married people live longer:

Selection Hypothesis: The selection hypothesis supposes that healthier people get more chances to get married. But most of the works did not support this statement and found only a very few exceptions.

Shared Lifestyle/Social Support Hypothesis: The most convincing explanation is that marriage creates a shared environment in which both spouses benefit from, for example, good nutrition, health insurance, and healthier behaviors. Married persons may also engage in fewer risky behaviours-for example, smoking or drinking less-behaviourally controlled by their spouse and family.

Shared Resources vs. Inequitable Resources:

Most past studies focused on the distinction in mortality between the married and unmarried populations but did not sufficiently test the shared risk idea.

Shared Resource Hypothesis: It postulates that when couples share income and other social support, they are benefited equally in terms of health and longevity.

Unequal Resource Hypothesis: Other researchers suggest that resources inside a marriage are not divided exactly equally. If the dominant spouse controls most of what is available in the marital household, the less resourceful spouse may face even higher risks of premature deaths due to relative deprivation.

Gaps in the Previously Conducted Research:

Mare and Pallon (1988): Most early work had relay on data, which lacks information about the better-resourced spouse. This limits the ability to test whether husbands and wifes experience similar or different mortality risks within the same marriage. Only a few studies, such as the work by Mare and Palloni (1988), have used detailed data on both partners to explore the shared risk of mortality. They examine how shared environment impacts the mortality experience of couples, exploiting the data in the National Longitudinal Survey of Mature Men (NLS). They found some evidence that some evidence that shared household factors such as household assets, affect the mortality risks of both spouses but very little evidence for the kind of broader shared effects captured by a couple-level unobserved factor.

Smith and McClean (1990): These authors investigated relationships between the deaths of husbands and wives by drawing on data from the National Health and Nutrition Examination Survey (NHANES) and its follow-up survey. They report that marriages between younger men and ones where at least one partner was a smoker carried the higher shared risk of dying.

However, their data did not provide detailed information about household environment, which prevented them from directly testing the shared lifestyle versus inequitable resource hypotheses.

Klein (1992): Klein studied over 1,100 married couples in the Framingham Heart Study and concluded that death times for the spouses were significantly correlated. Like previous studies, Klein assumed that there could only have been an unobserved frailty factor acting for both spouses, and his analysis did not have as a goal to estimate whether risk factors like smoking or wealth have different impacts on husbands compared to wives. Methods Overview

In this paper compared two of the primary theories over the influence of shared environment on the death of married partners:

Shared Lifestyle Hypothesis: Those who advance this theory suggest that it is the environmental factors-the bad diet or neighbourhood-and because both partners share that factor, then each wife and husband died or prospers to the equal extent of their spouse.

Inequitable Resource Hypothesis: This hypothesis predicts that the household resources such as income or social support may not be equally shared between the spouses. As a result, one spouse may be on the receiving end of the resources like higher income, while the other does not, which may have a potential to increase differences in mortality risks between the spouses.

These hypotheses inform an understanding of the connection between marital transitions

for example, divorce, and risks of mortality. According to the shared lifestyle hypothesis, divorced individuals may be at higher risk of mortality because of shared stress and loss of social or economic support. The inequitable resource hypothesis, on the other hand, suggests that effects of divorce may differ between husbands and wives because resource-poor spouses may have greater control over resources post-divorce, lowering their risk of mortality.

Estimation of the Model for Husbands and Wives Mortality

Classically, the problem of survival models with dependence of husbands' and wives' mortality rates was addressed through parameters that captured unobserved heterogeneity that reflects frailties shared between spouses.

However, the traditional estimation problem is encountered.

An alternative method, by Wei, Lin, and Weissfeld (1989), is to couple two hazard rate models, one for husbands and one for wives, through the correlation of their regression coefficients, not making assumptions about the dependence of event times within a couple. The model by Wei et al. assumes that: More than one failure time can occur in natural groupings, such as events that impact members of the same group.

The model employs information related to two individuals who are in the same relationship, contrasting their times or censoring with regards to death and covariates.

The model tests for shared environment versus inequitable resources by estimating the covariance matrix Cov(β) between the regression coefficients of husbands and wives. Estimation provides insights into whether covariates such as smoking or income generate similar or different mortality effects for husbands and wives.

This can be used to compare regression parameters across husbands and wives and determine whether certain covariates, such as smoking, have similar effects on both spouses' mortality risks. For example, if smoking leads to similar survival patterns in husbands and wives, the covariance of their regression parameters will be positive.

This dataset from the Panel Study of Income Dynamics (PSID) forms the basis for a survival analysis of married couples in the U.S. between 1968 and 1987. The purpose of this study is to determine mortality patterns based on socioeconomic, demographic, and behavioural factors, focusing specifically on shared household characteristics and their possible effects on the survival of husbands and wives.

Key Details of the Dataset:

Source: The PSID is a nationally representative sample of more than 5,000 U.S. households; the longitudinal data contain long periods on household dynamics.

Selection criteria:

Couples who had been married in 1968–1969 or during the first two years of PSID. There was an exclusion of women's prior marriages so the impact of previous marriages doesn't interfere with the conclusion.

Age of husband between 35 and 64 years in 1968; at this age husbands had likely completed their schoolings and childbearing. He is not old; thus, he has still not retired.

Longitudinal Data: The data are long series of households followed up until 1987, death, or nonresponse.

Sample Size: The last sample has 1,302 couples, and weights are used to correct for the initial sampling fractions and the time-varying nonresponse rates.

Covariates in Hazard Rate Models:

Time-Invariant Covariates: Social, economic, and demographic characteristics measured in 1968 or retrospectively in 1969 except for divorce data, which is time-varying.

Individual Risk Factors: The data set provides risk profiles including race, age, and disabilities.

Household Shared Resources: Sociodemographic characteristics, education, children ever born, household and county-level poverty, and health behaviours, such as smoking, risk avoidance.

Weighted Descriptive Statistics

Race: The sample is disproportionately white.

Age: Men have a mean age of 48 years, and their wives, 45 years.

Disability: An estimated 15% of men and 2% of women report activity-limiting disabilities.

Socioeconomic Profile: The average husband had about the same number of educational years as the average wife: roughly 11–12. Poor households were about 9 percent of all in 1968 or 1969; another 15 percent had been poor over both years. Divorced about 10 percent of couples by the period 1970-87.

Health Behaviours: Husbands are high risk avoiders. Household members averaged a collective consumption of about 18 cigarettes per day.

Follow-Up:

During the 17-year follow-up, 20% of the husbands died and 9% of the wives.

Summary of Outcomes from the Paired Hazard Rate Models

The paired hazard rate models examine the paired factors that influence the age-specific mortality risk of the couples. The models consider the initial age, disability, and husband's race that might influence the age at death. The results comprise two classes: individual factors and household factors influencing the spouse's mortality risk.

The initial individual risk factors related to the age-specific probability of death for the same period are:

Age: This should not be surprising because significantly increasing older age dramatically raises the mortality risk of both husbands and wives.

Disability: A key factor that raises the risk of mortality for both spouses is the presence of a disability.

Race: The effect of race on mortality risk was different for husbands and wives, though. The presence of a disability increased the mortality risk of wives with races other than white.

Household Environment Variables:

Cigarette Smoking: Mortality risk is significantly increased with more cigarettes smoked per day for husbands and wives, consistent with the shared resource hypothesis

Poverty: Household poverty in both 1968 and 1969 increases spouses' mortality risk; the effect is stronger when couples experience two consecutive years of poverty, consistent with shared resource hypothesis.

Number of Children: Incomes increase the chances of survival for both the husband and wife in having more kids. This is perhaps because parents become more motivated to become safer with the existence of children, and other children also provide them with more resources during old age.

Risk Avoidance: Lower risk behaviour in husbands measured by some sort of risk-avoidance score decreases the mortality, and there is some slight evidence that the wives experience the same phenomenon.

Education: Husband's Education: In this respect, the impact is virtually negligible on both sides.

Spousal Education: Increased educational levels among spouses decrease wives' potential risk of mortality and may represent spillover effects wherein the better-educated wives can embrace healthier lifestyles to the advantage of both spouses.

Divorce:

Husband: Divorce decreases the risk of mortality.

Wife: Divorce increases mortality risk and is a difference in the impact that divorce has on spouses, supporting the unequal resource hypothesis

Cross-Spouse Effects:

Age and Disability of Spouse: In the case of age, no strong proof of any cross-spouse effects seems to exist, and on the whole, individual disability does not have an adverse effect on spousal mortality.

Meaning of Shared Marital Environment:

Controlling for the fact that effects need to be bilateral for both spouses, there is evidence showing that

About 20% of the husbands' mortality risk is accounted for by the shared marital environment.

54% of the wives' mortality risk is accounted for by shared factors within the marriage, meaning that the shared marital environment has a greater effect on wives' mortality risk than on husbands.

Conclusion and Discussion

Studies in social epidemiology have emphasized how social factors, most importantly marriage, contribute to longevity. Nevertheless, what specific marital qualities have positive consequences for longevity are not well described. The present study addresses this gap by examining mortality risks among married couples, both at the individual and shared marital level. This study is different from other studies in that it uniquely compares mortality factors for both husbands and wives within a shared marital framework to explore whether these factors have consistent (shared resource hypothesis) or differing impacts (inequitable resource hypothesis) on each spouse's mortality risk.

What is remarkable about the sample used in the study is that it is very stable; it only includes couples married in 1968 or earlier, with each of them being in their first marriage and thus at a minimum age, making generalization to other cohorts or remarried couples an excellent avenue for further research. For middle-aged couples within this sample, the environmental contributions to mortality from smoking and household poverty are observable. Shared behaviours like smoking are missing detailed individual data, making it hard to distinguish active and passive effects, and so the smoking data of future studies should be more detailed. Poverty is another shared risk factor that likely represents deeper influences such as healthcare access, diet, and neighbourhood conditions that could further be enlightened with more detailed data.

Divorce is the uniquely influential factor which has differential effects on husbands and wives. After divorce, a man will lose social support, but he will gain financial security. A woman will have economic loss despite social network stability. This situation displays the fact that even if a man benefits financially after separation, the financial shocks for a woman may lead to stronger negative effects on her mortality risk. Overall, marriage encourages a shared environment in which both spouses, especially wives, share equal mortality risks, whereas the impact is shifted by divorce, thus indicating complexity in social and economic factors.

5.2 "The Effect of Widowhood on Mortality by the Causes of Death of Both Spouses" by Felix Elwert and Nicholas A. Christakis

Introduction

Widowhood has long been associated with increased mortality a phenomenon often referred to as the "widowhood effect." Studies have shown that losing a spouse can affect one's physical and mental well-being, putting the lives of widowed individuals at increased risk of mortality, particularly in the months following bereavement. However, most research done in this area addresses general all-cause mortality rather than specific causes of death. This kind of focus completely misses the subtleties with which widowhood influences mortality, especially whether the cause of death for the deceased spouse bears a relation to the risk of mortality of the surviving spouse. Felix Elwert and Nicholas Christakis argue in the paper, "The Effect of Widowhood on Mortality by the Causes of Death of Both Spouses," to close the knowledge gaps. In this study, they analysed whether widowhood affects cause-specific mortality, and the cause of death of the deceased spouse and its relationship with the survival partner's health outcomes.

This approach by this study provides an important dimension to bereavement research: the mechanism behind increased mortality and also the opportunity for targeted health care interventions can be detected if the widowhood effect is understood at a cause-specific level. Certain causes of death among the deceased spouse may have a relationship with higher risks of mortality in the surviving spouse; therefore, healthcare providers can use such information to make interventions for most vulnerable widowed individuals.

Objectives of the Study:

1.Is the risk of death among the surviving spouse increased due to the death of the spouse from a specific cause?

2.Does the cause of death of the deceased spouse modify the overall risk of death of the surviving spouse?

The authors had analysed their data separately for men and women because available evidence up to now indicates that widowhood affects both sexes differently. For instance, previous studies show that men suffer worse health declines after spouse loss and may be possibly due to diverging spousal loss coping strategies and utilization of social support. Trying to answer such questions related to the causes of widowhood effects based on the cause-specific approach, Elwert and Christakis delve into what it is specifically about health outcome that are particularly vulnerable to the effects of widowhood. Whether there is an independent role played by gender difference in the aspects remains under investigation.

In doing so, it draws from an enormous, nationally representative sample of elderly couples in the United States. The samples were based on Medicare files for 373,189 married couples who were monitored from 1993 until 2002. Being large in scale and consisting of a detailed demographic make this data quite reliable with which to work in conducting studies of widowhood and mortality. It consisted of older couples between 67 years and 98 years that lived in the same zip code and had enrolled themselves in Medicare. One exclusion has included only married Black and White couples and excluding enrolled participants in Health Maintenance Organizations (HMO). All these exclusions should tend to reduce racial differences in mortality rates that can be confounding.

Further categorization of cause-specific mortality into 16 broad categories was done by the authors: different cancers, cardiovascular diseases, Alzheimer's disease, and infection. It helped the authors to examine what kind of deaths are most associated with widowhood. Demographic and social covariates

such as race, age, socioeconomic status, and health history have also been used to control potential confounders that might lead towards an apparent result.

The authors applied the following statistical models:

1.Two competing risk models: That is, the impact of widowhood on the chances of death from a specified cause to the surviving spouse, accounting for dying from some other cause that is also a competing event.

2.Cox Proportional Hazards Models: This model estimated the effect of the widow's specific cause of death on the risk of survival of the spouse. It controlled for confounding variables and allowed the authors to isolate only the widowhood effect due to cause-specific mortality in the study.

These models offer a robust framework through which to understand the widowhood effect in a more nuanced manner. Accounting for confounding factors and focusing on specific causes, the analysis provides a detailed picture of how widowhood affects mortality in different health contexts.

Abstract: The study has been able to provide a clue on how widowhood influences mortality risk with significant differences observed according to the cause of death of the surviving spouse and the cause of death of the deceased spouse.

1. Higher Mortality Risk after Widowhood

The study was affirmative in confirming that indeed, widowhood increases the general mortality risk for both genders. For men, the risk of death increased by 18% after the loss of a wife, whereas for women, it was an increase of 16% after the loss of a husband. This increased risk is consistent with other literature on the widowhood effect, which shows that, generally, losing a spouse usually results in poor health conditions due to factors such as stress, loss of emotional support, and changes in health behaviours.

2. Variability in Cause-Specific Mortality

Importantly, the widowhood effect varied by cause of death. Spousal death substantially increased the risk of mortality of the surviving partner from specific causes. Both men and women had a heightened risk of dying from chronic obstructive pulmonary disease (COPD), infections or sepsis, diabetes, and lung cancer. Conditions that required ongoing monitoring and maintenance, like COPD and diabetes, seemed particularly impacted by widowhood. This may suggest that loss of a spouse interferes with the survivor's health maintenance and self-care behaviours or reduces the survivor's interest in health-related activities, thus exacerbating the condition's outcome in diseases that require constant care and attention.

3.Gender Differences in Mortality Risk

The findings revealed gender differences. Generally, men were more susceptible than women to die shortly after the loss of a partner, and mortality causes ranged from COPD to infectious diseases and lung cancer in particular. Perhaps such trends are an indirect indication that men are typically more dependant on the wife who provides support at all kinds of levels while maintaining every day's activity as well as health management. Accordingly, the widowhood itself might put the men in that position at more risk for some range of health-related issues; and it is assigned to a sudden loss of an important figure from whom one could count upon for proper care or even support.

4.Cause of Death from Which the Partner Died

It was found that there had been an influence by cause of death of which the partner had died in affecting the death of the surviving partner. For example, if a wife had died of lung cancer, infections, or COPD, husbands had a significantly elevated risk of dying. That may be because the couple shares common health behaviours or exposures (such as smoking) that contributed to their risks. Death due to Alzheimer's or Parkinson's disease in a spouse didn't increase the risk of mortality for the surviving spouse. The authors suggest that this might be due to "anticipatory grief," given that the surviving spouse is able to prepare emotionally and practically for the gradual demise of their partner.

Discussion and Implications

The study's findings underscore the complexity of the widowhood effect, suggesting that the health impacts of bereavement vary significantly based on both the cause of death and gender. Widowhood seems to affect mortality through a range of mechanisms, including biological, psychological, and social pathways. The variation in mortality by specific causes suggests that widowhood may disrupt social and behavioural supports that are crucial for managing chronic illnesses.

Besides, it highlights the importance of gender in the comprehension of health effects of widowhood. There is a gender effect for men and women such that the risks for widowhood seem to be more elevated in men, but for most conditions, significantly higher, especially those with links to lifestyle factors and patient management. This will go on to inform healthcare policy and social support strategies tailored toward providing gender-sensitive support to widowed individuals.

Targeted health interventions, such as counselling, health monitoring, and community support, could mitigate the widowhood effect for chronic diseases that require active management. Healthcare providers might reduce mortality risk in this vulnerable population by focusing on people who have recently lost a spouse. The results of this research would also be useful for planning social support structures directed specifically to the needs of widows, particularly in instances of complex health.

Limitations

The research study is methodologically robust with some limitations. Since reliance on Medicare records can draw inferences of cause of death, it could likely under-report sudden and other undiagnosed causes that may be present as evident in death certificates. In addition, since the study controlled for a lot of variables, there might still be an effect of unmeasured lifestyle factors, such as smoking or diet, especially on conditions like COPD or lung cancer. This study only included White and Black people, which reduces the generalizability of the findings to other racial and ethnic groups who may have different cultural practices or social support systems.

Conclusion

This research work carried out by Elwert and Christakis does give the analysis of cause-specific mortality because of widowhood to prove that it leads the way for various risks toward death, particularly among cases related to chronic management and acute health events. Mortality risks through widowhood were seen significant by research, and this, furthered by the studies, presents that social support along with health interventions might critically address such a risk factor in terms of reducing death or the rate of deaths over a period.

This study has very important implications for public health and geriatric care, especially when designing interventions specifically targeted for recently widowed people. Further studies may build on this work by studying the widowhood effect in various populations and analyzing biological, psychological, and behavioural mechanisms that underlie this effect. As populations age, so will understanding and addressing the health challenges related to widowhood be increasingly relevant for healthcare providers and policymakers.

Discussion and Conclusion

The study reveals a significant interdependence between the mortalities of husbands and wives. It highlights the social issue of surviving the loss of a spouse and how this affects the remaining partner's longevity. Our comparative analysis demonstrates that the probability of survival for wives is generally higher than that for husbands after the loss of a spouse. As estimated life expectancy of remaining life of wife after death of husband is 20.5 years and estimated life expectancy of remaining life of of husband after death of wife is 16 years.

This study investigated the impact of spousal mortality on the survival probability of the surviving partner, using survival analysis techniques. Our findings indicate a significant interdependence between spousal mortality rates, consistent with previous research by Smith and Zick (2006) and Elwert and Christakis (2008). The Kaplan-Meier survival curves and Cox proportional hazards models and Copula model revealed that the death of one spouse substantially increases the mortality risk for the surviving partner.

These results have important implications for the design of joint life insurance products and support services for surviving spouses. By understanding the factors that influence spousal survival, policymakers and insurance companies can develop more effective interventions to support individuals during these critical times. These findings offer crucial insights for developing joint life insurance policies within the insurance sector, emphasizing the need for tailored support and services to address the unique challenges faced by the surviving spouse.Future research should explore annuities regarding joint life, lifetables etc.

Appendix

Python Code 1 (Survival curves for X1 and X2)

```
import pandas as pd
from lifelines import KaplanMeierFitter
from lifelines.statistics import logrank_test
import matplotlib.pyplot as plt
import seaborn as sns
df=pd.read_excel("E:/Project/350 obs 1.xlsx")# Data preparation
df = pd.DataFrame(df)
kmf = KaplanMeierFitter()
kmf.fit(df['X1'], event_observed=df['SX1'])
kmf.plot_survival_function()
plt.title('Survival Function for Husband')
plt.show()
```

Python Code 2

```
# Kaplan-Meier Fitter (Wife)
kmf.fit(df['X2'], event_observed=df['SX2'])
kmf.plot_survival_function()
plt.title('Survival Function for Wife')
plt.show()
```

Python Code 3 (Survival curves for X1 and X2)

```
sns.histplot(df['X3'], bins=5, color='blue', kde=True, alpha=0.5)
plt.title('Histogram for Remaining Age of Husband after Death of Wife')
plt.xlabel('Ages')
plt.ylabel('No of Husbands')
plt.show()
```

Python Code 4

```
import seaborn as sns
import matplotlib.pyplot as plt

sns.histplot(df['X4'], bins=5, color='red', kde=True, alpha=0.5)

plt.title('Histogram for Remaining Age of Wife after Death of Husband')
plt.xlabel('Ages')
plt.ylabel('No. of Wives')
plt.show()
```

Python Code 5 (Survival plot of X3 and X4)

```
df = df.dropna(subset=['X3', 'SX3'])
kmf = KaplanMeierFitter()
kmf.fit(df['X3'], event_observed=df['SX3'])

kmf.plot_survival_function()
plt.xlim(0, 50)

# Customize the plot
plt.title('Survival Function Plot')
plt.xlabel('Time (Ages of Husband)')
plt.ylabel('Survival Probability')
plt.show()
df = df.dropna(subset=['X4', 'SX4'])

kmf = KaplanMeierFitter()
kmf.fit(df['X4'], event_observed=df['SX4'])

kmf.plot_survival_function()
plt.xlim(0, 50)

plt.xlabel('Time (Ages)', labelpad=5)
plt.ylabel('Survival Probability')
plt.title('Survival Function Plot for X4')
plt.subplots_adjust(left=0.1, right=0.9, top=0.9, bottom=0.2)
plt.show()
```

Python Code for Kaplan-Meier Survival Function (X5 and X6)

```
# Drop rows with missing values in 'X5' or 'SX5'
df_clean = df.dropna(subset=['X5', 'SX5'])

# Fit the Kaplan-Meier model using the cleaned data
kmf = KaplanMeierFitter()
kmf.fit(df_clean['X5'], event_observed=df_clean['SX5'])

# Plot the survival function
plt.figure(figsize=(10, 6))
kmf.plot_survival_function(color='red')

plt.xlim(20, 90)
plt.xlabel('Time (Ages)', labelpad=5)
plt.ylabel('Survival Probability')
plt.title('Survival Function Plot for X5')
plt.subplots_adjust(left=0.1, right=0.9, top=0.9, bottom=0.2)
plt.grid(True)
```

```
plt.show()

df_clean = df.dropna(subset=['X6', 'SX6'])

kmf = KaplanMeierFitter()
kmf.fit(df_clean['X6'], event_observed=df_clean['SX6'])

plt.figure(figsize=(10, 6))
kmf.plot_survival_function(color='green')

plt.xlim(20, 80)
plt.xlabel('Time (Ages)', labelpad=5)
plt.ylabel('Survival Probability')
plt.title('Survival Function Plot for X6')
plt.subplots_adjust(left=0.1, right=0.9, top=0.9, bottom=0.2)
plt.grid(True)
plt.show()
```

```
import matplotlib.pyplot as plt
from lifelines import KaplanMeierFitter
df_filtered = df[(df['SX1'] == 1) & (df['SX2'] == 1)]

kmf_husband = KaplanMeierFitter()
kmf_wife = KaplanMeierFitter()

kmf_husband.fit(df_filtered['X1'], event_observed=df_filtered['SX1'
    ], label='Remaining age of wife after death of husband')
kmf_wife.fit(df_filtered['X2'], event_observed=df_filtered['SX2'],
    label='Remaining age of husband after death of wife')

plt.figure(figsize=(10, 6))
kmf_husband.plot_survival_function(color='blue')
kmf_wife.plot_survival_function(color='green')

plt.xlim(0, 90)
plt.title('Survival Function Plot: Husband vs Wife Ages ')
plt.xlabel('Time (Ages)')
plt.ylabel('Survival Probability')
plt.legend()
plt.grid(True)
plt.show()
```

Python Code for Kaplan-Meier Survival Curves (Husband vs Wife)

```
import matplotlib.pyplot as plt
from lifelines import KaplanMeierFitter
```

```

kmf_husband = KaplanMeierFitter()
kmf_wife = KaplanMeierFitter()

fig, axes = plt.subplots(1, 2, figsize=(14, 6))  # 1 row, 2 columns
   of plots
for subset_label in df['subset_H'].unique():
    mask = df['subset_H'] == subset_label

    # Fit Kaplan-Meier for husband's data
    kmf_husband.fit(df['X1'][mask], event_observed=df['SX1'][mask],
        label=f'Husband - {subset_label}')
    kmf_husband.plot_survival_function(ax=axes[0])  # Plot on the
        left subplot

# Set title and labels for husband's plot
axes[0].set_title('Husband Survival Curves for Different Subsets')
axes[0].set_xlabel('Time (Age)')
axes[0].set_ylabel('Survival Probability')
axes[0].set_xlim(25,90)
axes[0].legend(title='Subset')
axes[0].grid(True)

for subset_label in df['subset_W'].unique():
    mask = df['subset_W'] == subset_label

    kmf_wife.fit(df['X2'][mask], event_observed=df['SX2'][mask],
        label=f'Wife - {subset_label}')
    kmf_wife.plot_survival_function(ax=axes[1])

axes[1].set_title('Wife Survival Curves for Different Subsets')
axes[1].set_xlabel('Time (Age)')
axes[1].set_ylabel('Survival Probability')
axes[1].set_xlim(25,90)
axes[1].legend(title='Subset')
axes[1].grid(True)

# Adjust layout to avoid overlap
plt.tight_layout()
plt.show()
```

Python Code for Log-Rank Test Comparison (Husband vs Wife)

```
# Let's compare '1920-50' vs '1950-1980'
group_A = df[df['subset_H'] == '1920-55']
group_B = df[df['subset_H'] == '1955-90']
```

```

# Apply the log-rank test
result = logrank_test(group_A['X1'], group_B['X1'], event_observed_A
    =group_A['SX1'], event_observed_B=group_B['SX1'])

print(result)

# Let's compare '1920-50' vs '1950-1980'
group_A = df[df['subset_W'] == '1920-55']
group_B = df[df['subset_W'] == '1955-90']

result = logrank_test(group_A['X2'], group_B['X2'], event_observed_A
    =group_A['SX2'], event_observed_B=group_B['SX2'])

print(result)
```

Histogram of Age difference

```
plt.figure(figsize=(8, 6))
sns.histplot(df['age_difference'], kde=True, bins=10)
plt.title("Distribution␣of␣Age␣Differences␣between␣Husband␣and␣Wife"
    )
plt.xlabel("Age␣Difference␣(Husband␣-␣Wife)")
plt.ylabel("Frequency")
plt.show()
```

R Codes

R Code 1

```
library(ggplot2)
library(survminer)
data = read_excel("D:/vivek/Sem3/Research Project/350.xlsx
    .")
attach(data)
surv_obj1 = Surv(time = data$X1, event = as.numeric(data$
    SX1))
fit1 = survfit(surv_obj1 ~ 1, data = data)
summary(fit1)
```

R Code 2

```r
surv_obj2 = Surv(time = data$X2, event = as.numeric(data$
   SX2))
fit1 = survfit(surv_obj2 ~ 1, data = data)
summary(fit2)
```

R Code 3

```r
combined_fit1=survfit(Surv(time, event) ~ group, data =
   combined_data)
#plot the survival curves
plot_combined=ggsurvplot(combined_fit1,
        title = 'Kaplan-Meier Survival Curves for Husband
           and Wives(Complete Ages)',
        xlab='Time',
        ylab='Survival Probability',
        legend.title='Group',
        lengend.labs=c('Husband','Wives'),
        risk.table=TRUE);plot_combined
```

R Code 4

```r
#create survival objects
surv_obj1=Surv(time=data$X1,event=as.numeric(data$SX1))
surv_obj2=Surv(time=data$X2,event=as.numeric(data$SX2))
surv_obj3=Surv(time=data$X3,event=as.numeric(data$SX3))
data_pair1=data.frame(time=c(data$X1, data$X2),event=c(as.
   numeric(data$X1),as.numeric(data$X2)),
group = factor(rep(c('X1','X2'), each= length(data$X1)))
data_pair2=data.frame(time=c(data$X3, data$X4),event=c(as.
   numeric(data$X3),as.numeric(data$X4)),
group = factor(rep(c('X3','X4'), each= length(data$X3)))
log_rank_test_1_2=survdiff(Surv(time,event) ~ group, data
   = data_pair1)
log_rank_test_3_4=survdiff(Surv(time, event) ~ group, data
    = data_pair2)
print(log_rank_test_1_2
print(log_rank_test_3_4
```

R code 5

```r
data_pair2=data.frame(time=c(data$X3, data$X4),event = c(
   as.numeric(data$SX3), as.numeric(data$SX4)), group =
   factor(rep(c('X3','X4'), each = length(data$X3)))
cox_fit_oair2=coxph(Surv(time, event) ~ group, data = data
   _pair2)
summary(cox_fit_pair2)
```

R code 6

```r
combined_data=data.frame(time = combined_time2,event=as.
   numeric(combined_event2), group=combined_group)
combined_fit2=survfit(Surv(time, event) ~ group, data=
   combined_data)
plot_combined2=ggsurvplot(combined_fit2,title='kaplan-
   meier survival curves for Husbands and Wives(Remaininig
    Ages)',xlab='Time',ylab='Survival Probability', legend
    .title='Group', legend.labs=c('Husband','Wives'), risk.
   table=TRUE);plot_combined2
```

```r
data_pair1 <- data.frame(time = c(data$X1, data$X2), event
    =c(as.numeric(data$SX1), as.numeric(data$SX2)),group =
    factor(rep(c("X1", "X2"), each = length(data$X1))))
# Apply Cox Proportional Hazards Model
cox_fit_pair1 <- coxph(Surv(time, event) ~ group, data =
   data_pair1)
summary(cox_fit_pair1)
```

```r

# Define a function to calculate life expectancy from a
   survival fit
calculate_life_expectancy <- function(surv_fit) {
  surv_summary <- summary(surv_fit)
  survival_probs <- surv_summary$surv
  time_points <- surv_summary$time
  life_expectancy <- sum(survival_probs * diff(c(0, time_
     points)))
  return(life_expectancy)
}
```

References:

- Smith, K. R., & Zick, C. D. (2006). Linked Lives, Dependent Demise? Survival Analysis of Husbands and Wives. *Journal of Demographic Studies, 23*(4), 450-467.
- Elwert, F., & Christakis, N. A. (2008). The Effect of Widowhood on Mortality by the Causes of Death of Both Spouses. *American Journal of Public Health, 98*(11), 2092-2098.
- Wang, P., Li, Y., & Reddy, C. K. (2019). Machine Learning for Survival Analysis: A Survey. *ACM Computing Surveys, 51*(6), 1-36.
- Kleinbaum, D. G., & Klein, M. (2012). *Survival Analysis: A Self-Learning Text, Third Edition*. Springer.
- Charpentier, A. (2015). *Computational Actuarial Science with R*. Routledge.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *Introduction to Statistical Learning, Second Edition*. Springer.

Software Used: Excel, R, Python, Minitab.