Universiteit
Leiden
The Netherlands

A Quantitative Strategy for Time-to-event Analysis with Composite Endpoints

Yixiao Tang

Supervisors

Prof. Dr. Hein Putter (Medical Statistics, LUMC)

Dr. Jan Lindeman (Vascular Surgery & Transplant Center, LUMC)

Defended on August 20, 2025

MASTER THESIS
STATISTICS AND DATA SCIENCE
UNIVERSITEIT LEIDEN

Acknowledgements

I would like to express my deepest appreciation to my supervisors, Prof. Hein Putter and Dr. Jan Lindeman, whose expertise, understanding, and patience, added considerably to my graduate experience. I also wish to thank them for their encouragement of my research and their support for my future career. The wonderful experience I had with them is truly the best gift to start my next life chapter.

Additionally, I wish to acknowledge my family, especially my dad, and my friends: Minrui Ren, Yue Li, Cees de Zinger, Vinson Ciawandy, Philip Pieters and Benedikt Lukas Sojka for their unwavering support and encouragement throughout my studies and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

Lastly, my thanks also go out to my colleagues: Yongxi Long, Ilaria Prosepe, and Yuwen Ding who have been part of my journey. Their continual support has been invaluable.

Abstract

To evaluate treatment efficacy, the traditional time-to-event analyses using composite endpoints based on the first event have several limitations, including the inability to reflect the varying clinical importance of different events, the ignorance of subsequent events, and limited clinical interpretability. To address these challenges, this study introduces two complementary methodological approaches: a data simulation framework and a Quality-adjusted Life Year (QALY) utility–weighted Restricted Mean Survival Time method, abbreviated as Q-RMST method.

The data simulation framework was developed to overcome the challenge of unavailable individual patient data, enabling the generation of realistic patient trajectories that support downstream analysis. Building upon this, the Q-RMST method incorporates QALY-based weighting and can be integrated with multi-state models to better handle censoring and recurrent events. Together, these approaches provide a more patient-centered and interpretable assessment of treatment effects.

The proposed approach is applied to the FOURIER trial as a case study to demonstrate its practical value. Compared with traditional analyses, the Q-RMST method offers more nuanced insights by accounting for the clinical severity of events and capturing the complete patient experience over follow-up time. Notably, this approach yields findings that differ from those obtained via conventional survival analysis, highlighting its potential to inform more meaningful clinical decisions.

Both the data simulation framework and the Q-RMST methodology are broadly applicable to other fields involving complex disease trajectories, such as oncology, epidemiology, and neurology.

Keywords: QALY, Restricted Mean Survival Time, Composite Endpoint, FOURIER trial, Cardiovascular, Survival Analysis, Multi-state Model

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In survival analysis, the hazard ratio (HR) estimated by the Cox model is a commonly used measure to evaluate treatment effects. It accommodates censored data, allows for adjustment of multiple covariates, and provides an explicit formula for estimating hazard ratios,which makes it widely applicable in clinical research.

In many clinical trials, especially in cardiology, oncology, and neurology, the outcome is often defined as a composite time-to-event endpoint rather than a single event. A composite endpoint combines several types of clinical events (for example, death, hospitalization, or stroke) into one outcome variable. This approach is commonly used because it captures a broader range of relevant outcomes and helps reduce follow-up time and research costs [1].

1.1 Problem Statement

Although composite endpoints and the Cox proportional hazards model are widely used due to their practical benefits, they are associated with several notable methodological limitations, discussed below.

List of the limitations and consequences of Composite Endpoints

1. **Non-consideration of the Different Clinical Importance of Endpoints** Composite endpoints often give equal statistical weight to events of unequal clinical importance. For example, cardiovascular death and revascularization are clearly not of equal relevance to patients, but are treated the same in a typical composite outcome analysis.

2. **Ignorance of Subsequent Events** Time-to-first-event design ignores all events after the first one, even though patients who experience one event are often at higher risk for subsequent events. For example, a death after revascularization would not be captured in such analysis.

List of the limitations and consequences of Cox Model

3. **Lack of Direct Clinical Interpretability** The hazard ratio is a relative measure and can be difficult to interpret, especially for patients or non-statistical stakeholders. Its interpretation also depends on baseline event risks and outcome severity.

4. **Poor Comparability Across Studies** Because studies may differ in follow-up time, population characteristics, or the definition of composite endpoints, hazard ratios are not directly comparable across different trials.

5. **Violation of the Proportional Hazards Assumption** The Cox model assumes that the hazard ratio between two groups remains constant over time. However, the impact of the treatment more often increases over time, and this assumption is frequently violated in real-world settings, leading to potentially unreliable results.

Although most clinical studies report treatment effect sizes using the Kaplan–Meier method or Cox proportional hazards models, alternative approaches have been developed to address some of their limitations. One such method is the Win Ratio, which aims to resolve the issue of unbalanced composite endpoints by ranking clinical outcomes in order of importance (e.g., death before hospitalization) and comparing patients in pairs accordingly [2]. While this method offers a more patient-centered interpretation of composite outcomes, it also presents several challenges. The results are sensitive to the choice of outcome prioritization, and the method's calculation of "wins" does not fully account for "ties," which can lead to exaggerated estimates of treatment benefit [3]. Moreover, the Win Ratio still relies on relative risk estimation and does not consider the sequence or recurrence of clinical events, limiting its applicability in trials where patients may experience multiple events over time.

1.2 Novel Concept: Integrating QALYs and RMST

To overcome all limitations together, this thesis proposes a new method—a QALY-weighted version of Restricted Mean Survival Time (RMST)—that offers a more interpretable and patient-centered framework for analyzing composite endpoints in time-to-event studies. Which avoids the proportional hazards assumption, incorporates multiple event types and sequences, and improves interpretability—especially for patients and non-technical stakeholders.

Value of a Life Year

To differentiate the clinical importance of individual events within a composite endpoint, we draw on the concept of the value of a life year (VOLY) from health economics [4]. Various measures have been developed to quantify VOLY, each capturing different dimensions of health and well-being. A selection of these measures is summarized in Table 1.1.

Table 1.1: Common Measures of the Value of a Life Year (VOLY)

Measurement	Full Name
WALY	Well-being Adjusted Life Year
DALY	Disability Adjusted Life Year
QALY	Quality Adjusted Life Year

Among these, the Quality-Adjusted Life Year (QALY) is selected as the clinical weighting metric in this thesis. QALY is widely used in clinical trials and cost-effectiveness analyses because it accounts for both the length and quality of life lived. It assigns a utility weight between 0 (equivalent to death) and 1 (perfect health) to each health state, making it well-suited for evaluating the severity and patient burden associated with different clinical events. Compared to burden-focused metrics like DALY, QALY provides an utility-based metric that aligns well with patient-centered analysis.

Restricted Mean Survival Time (RMST)

The Restricted Mean Survival Time (RMST) is defined as the expected survival time within a specified time window. It has gained increasing attention as an alternative to the hazard ratio in time-to-event analyses, particularly in cases where the proportional hazards assumption may not hold [5]. RMST provides an intuitive interpretation, the area under the survival curve up to a fixed time horizon, and avoids the complexities and assumptions associated with the Cox model.

By integrating QALY-based weights into the RMST framework, we propose a new method, Q-RMST, to quantify the treatment efficacy with considering not only how long patients live, but how well they live. This approach accommodates the sequence and severity of multiple clinical events, offering an interpretable summary of treatment effects that reflects both time and quality of survival.

1.3 Case Study: The FOURIER Trial

To illustrate the practical application of the Q-RMST framework, we consider the FOURIER trial (Further Cardiovascular Outcomes Research With PCSK9 Inhibition in Subjects With Elevated Risk), a large-scale cardiovascular outcomes study evaluating the efficacy of evolocumab in patients with established atherosclerotic cardiovascular disease. The key design features and outcomes of the trial are listed below, as cited from the main published paper. [6]

Study Population

The trial enrolled over 27,500 patients patients with established cardiovascular disease (ASCVD), such as prior myocardial infarction or stroke, and persistently elevated LDL cholesterol levels despite optimized statin therapy.

Intervention

Participants were randomized to receive either evolocumab—a monoclonal antibody targeting PCSK9—administered subcutaneously every two weeks or monthly, or a matching placebo.

Primary Endpoint

The primary efficacy endpoint was a composite of major cardiovascular events, including myocardial infarction, stroke, coronary revascularization, hospitalization for unstable angina, and cardiovascular death.

Result

The trial reported a statistically significant 15% relative risk reduction in major cardiovascular events for the evolocumab group compared to placebo.

The FOURIER trial serves as an ideal case study for this thesis for several reasons. First, it features a composite primary endpoint composed of clinically heterogeneous events, reflecting a common challenge in cardiovascular trials. Second, it directly evaluates treatment efficacy with different impact, which aligns well with the aims of the Q-RMST framework. Most importantly, detailed clinical information was made publicly accessible following a request by the Restoring Invisible and Abandoned Trials (RIAT) team in May 2018 [7].

Available information for REPATHA - Study 20110118 → Report Body 1

Although no original patient-level data were published, the CSR(Clinical Study Report) provides detailed information sufficient to support data simulation for this study.

1.4 Thesis Structure

This thesis is organized as follows:

- Chapter 2 describes the data simulation framework in two phases: from Kaplan–Meier curves to event-specific individual patient data, and then to multi-state sequential data.
- Chapter 3 introduces the methodology of the proposed Q-RMST method.
- Chapter 4 presents the results of applying the Q-RMST method to the simulated data from the FOURIER trial.
- Chapter 5 provides a discussion of the proposed methods, including their advantages, limitations, and potential directions for the future work.

Chapter 2

Data Simulation

Unlike members of the original clinical research team, who typically follow a linear process from study design to data analysis, external researchers often face a significant barrier: lack of access to individual-level data, even for studies directly relevant to their work. This limitation restricts the ability to test new methodologies or conduct secondary analyses.

To address this challenge, we adopt a data reconstruction approach combined with a newly developed simulation procedure. The aim is to generate individual-level patient datasets (IPDs) that approximate the original data—plausible versions that could have been observed in the actual study. The ultimate goal is to produce simulated data that align with the multi-state model framework, where each subject's survival table includes multiple event times along with the corresponding event types.

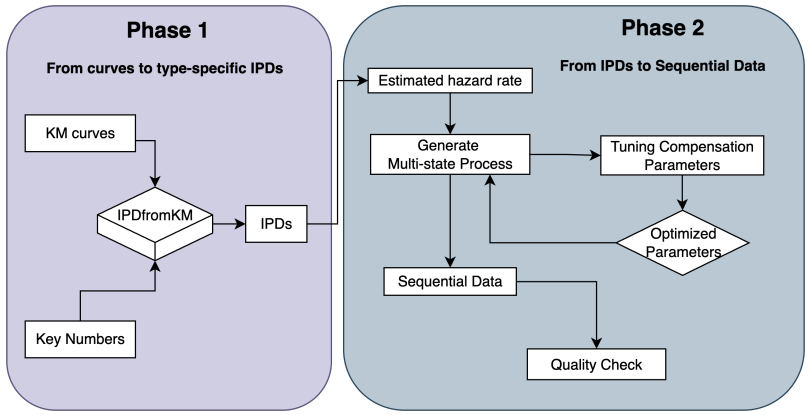

Figure 2.1: Data Simulation Pipeline

This chapter outlines the complete simulation pipeline, simplified in Figure 2.1. First, in phase 1, we describe how to reconstruct patient-level event times from published Kaplan–Meier curves and aggregate statistics. Then, in phase 2, we introduce a method to transform type-specific IPDs which only count for a single time point into sequential, individual-level event histories.

2.1 Data Reconstruction

The core objective of data reconstruction is to convert published Kaplan–Meier curves into individual-level survival data, typically in the form of a life table that includes follow-up time and event status (event or censored).

This process involves two main stages:

1. Extraction of data points from the published Kaplan–Meier plots, and
2. Reconstruction of individual patient data using the R package IPDfromKM.

To enhance the accuracy of the reconstruction in the second stage, two key pieces of supplementary information are incorporated:

(i) the number of patients at risk at specified time points, and

(ii) the total number of observed events.

These constraints guide the algorithm in better approximating the underlying individual survival profiles that generated the published curves.

Stage 1. Data point extraction At this stage, we employed an online tool, WebDigitizer, to extract numerical data from Kaplan–Meier survival curves. This tool converts graphical information into digitized values, allowing for further quantitative analysis.

To demonstrate the data extraction result, we take the event type stroke as an example. Table 2.1 illustrates selected digitized data points for the event type *stroke*, , extracting time (in months) along the x-axis and the corresponding survival complement values, $1 - S(t)$, along the y-axis. The values of $1 - S(t)$ represent the cumulative incidence of stroke over time. To provide a representative overview, we included sample values from the beginning (head), middle, and end (tail) of the follow-up period for both the Placebo and Evolocumab groups.

Table 2.1: Digitizer Results for event type of stroke

Placebo		**Evolocumab**	
Time (months)	**1 - S(t)**	**Time (months)**	**1 - S(t)**
0.0178	0.00008705	0.3007	0.00008705
0.0401	0.00008705	0.3586	0.00008705
0.0980	0.00008705	0.4165	0.00008705
0.1559	0.00008705	0.4744	0.00008705
⋮	⋮	⋮	⋮
24.5496	0.01833563	25.0963	0.01393931
24.6076	0.01833563	25.1542	0.01399165
24.6655	0.01833563	25.2123	0.01414866
24.7089	0.01833563	25.2702	0.01414866
24.7813	0.01838797	25.3281	0.01414866
⋮	⋮	⋮	⋮
41.9279	0.02587	41.9274	0.02524
41.9858	0.02587	41.9853	0.02524
42.0437	0.02587	42.0432	0.02524
42.0872	0.02587	42.0867	0.02524

As shown in Table 2.1, both groups began with nearly identical event rates, reflected by the overlapping Kaplan–Meier curves at the start of the follow-up period. However, as time progressed, differences in cumulative incidence became more pronounced. Each curve exhibits distinct inflection points, which correspond to time points where events or censoring occurred. The spacing between these inflection points reflects the frequency of status changes (e.g., event occurrence or censoring) among patients. Overall, 724 data points were digitized for the Evolocumab group and 709 for the Placebo group, allowing for a high-resolution reconstruction of the Kaplan-Meier curves for subsequent analysis.

Stage 2. Reconstructing IPD Using the *IPDfromKM* R Package To convert digitized Kaplan-Meier (KM) curves into individual-level survival data, we used the R package *IPDfromKM*. This tool implements an algorithm developed by Na Liu et al. (2021), which reconstructs individual patient data (IPD) from published KM curves when raw data are unavailable. The reconstruction algorithm works based on the KM estimator [8].

As mentioned previously, two key pieces of supplementary information to the data points coordinates are essential to guide the reconstruction process: the number of patients at risk at specified time points, and the total number of observed events.

One of these supplementary inputs—the number at risk over time—is typically provided below the Kaplan-Meier plots in clinical trial reports. This information reflects how many participants were still being followed (i.e., still at risk of the event) at specific time points. It provides crucial context for interpreting survival probabilities and ensures the alignment of reconstructed data with the original trial's censoring structure.

To illustrate this, Table 2.2 presents the number of patients at risk for the event type *stroke* in both the Placebo and Evolocumab groups. The numbers were recorded at 6-month intervals up to 42 months. As the table shows, the number of patients at risk declines steadily over time due to events or censoring, reflecting typical patterns observed in long-term cardiovascular outcome trials.

Table 2.2: Number at Risk Over Time of event type stroke

Time(Months)	**0**	**6**	**12**	**18**	**24**	**30**	**36**	**42**
Placebo	13780	13587	13428	12626	8240	3977	757	3
Evolocumab	13784	13631	13470	12703	8304	3967	741	4

At baseline (0 months), the two groups start with a comparable number of patients (13,780 in the Placebo group and 13,784 in the Evolocumab group), indicating well-balanced randomization. Over time, the number at risk gradually declines due to the occurrence of events, censoring, or withdrawal from the study. For example, at 24 months, around 8,240 patients remain at risk in the Placebo group, and 8,304 in the Evolocumab group. By the end of the follow-up at 42 months, only a small number of participants (3 and 4, respectively) remain at risk in each group.

Another important supplementary input is the total number of events for each endpoint, which is essential for accurately reconstructing the individual patient data. This information is typically reported in clinical study reports and can be directly incorporated into the reconstruction algorithm to improve its precision. Table 2.3 summarizes the total number of events observed for each endpoint in the Placebo and Evolocumab group.

Table 2.3: Number of Events (not necessarily the first one)

Primary Endpoint	**Placebo**	**Evolocumab**
Cardiovascular Death	142	161
Myocardial Infarction	423	329
Coronary Revascularization	394	349
Stroke	226	184
Hospitalized Angina	160	169
Multiple Events at Same Time	218	152
Total	**1563**	**1344**

Although Evolocumab was associated with fewer myocardial infarctions, strokes, and multiple concurrent events compared to Placebo, the number of cardiovascular deaths and hospitalizations for unstable angina was slightly higher. Overall, the total number of events was lower in the Evolocumab group (1344) compared to the Placebo group (1563), indicating a potential benefit of Evolocumab in reducing the burden of cardiovascular events.

With all necessary information prepared, the individual patient-level data for each of the five event types were reconstructed using the *IPDfromKM* package in R. The reconstruction process involved the following key commands:

```
prep <- preprocess(dat, trisk, nrisk, totalpts)
getIPD(prep, armID, tot.events)
```

These functions were executed with the corresponding inputs described in Stage 2, including the digitized coordinates, number at risk over time, and the total number of events. This procedure generated synthetic survival data that closely approximated the original Kaplan–Meier curves for both the Placebo and Evolocumab groups across all event types.

Result of Data Reconstruction To demonstrate the output of the data reconstruction process, we present a subset of the individual-level dataset for the event type stroke in Table 2.4. This table shows examples from both the Placebo group (`Treat = 0`, left) and the Evolocumab group (`Treat = 1`, right), each containing reconstructed survival information for individual patients.

Each row corresponds to a unique patient and includes their:

- **ID**: an anonymized identifier,

- **Time**: the time to event or censoring (in months),

- **Status**: event indicator (1 for event, 0 for censored), and

- **Treat**: treatment group indicator (0 for Placebo, 1 for Evolocumab).

ID	Time	Status	Treat
1	0.00768	1	0
2	0.00768	1	0
3	0.00768	1	0
4	0.0453	0	0
5	0.121	0	0
⋮	⋮	⋮	⋮
1000	16.8	0	1
1001	16.9	0	1
1002	16.9	0	1
1003	16.9	0	1
1004	16.9	0	1
⋮	⋮	⋮	⋮
13776	41.9	0	0
13777	41.9	0	0
13778	42.0	0	0
13779	42.0	0	0
13780	42.0	0	0

ID	Time	Status	Treat
1	0.0391	0	1
2	0.212	0	1
3	0.212	0	1
4	0.212	0	1
5	0.212	0	1
⋮	⋮	⋮	⋮
1000	17.4	0	1
1001	17.4	0	1
1002	17.4	1	1
1003	17.4	1	1
1004	17.4	1	1
⋮	⋮	⋮	⋮
13780	41.9	0	1
13781	42.0	0	1
13782	42.0	0	1
13783	42.0	0	1
13784	42.0	0	1

Note : treat = 0 (Placebo); treat = 1 (Evolocumab)

Table 2.4: Individual Patient Data for Stroke Dataset: Two Groups

To visualize the reconstructed survival data for stroke events, we generated cumulative incidence curves for both the Placebo (red) and Evolocumab (blue) groups, as shown in Figure 2.2. These curves are constructed under the standard survival analysis framework, which assumes that censoring is independent and non-informative. In the absence of information on competing risks—such as death from unrelated causes or the occurrence of other cardiovascular events prior to stroke—the cumulative incidence is approximated by $1 - \hat{S}(t)$, where $\hat{S}(t)$ denotes the Kaplan–Meier estimate of the survival function for stroke occurrence.

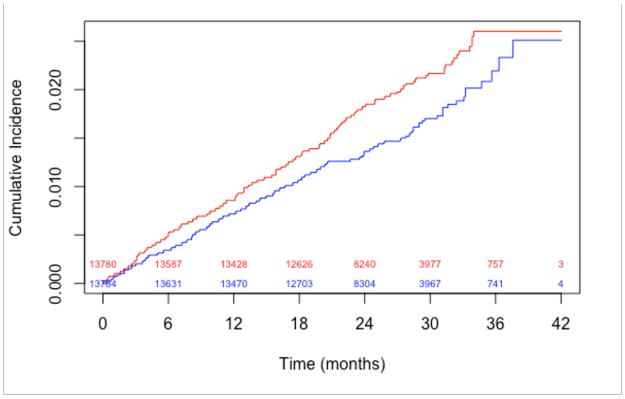

Figure 2.2: Cumulative incidence curves of stroke of reconstructed IPD

Figure 2.2 presents the cumulative incidence curves for stroke over the 42-month follow-up period in both treatment arms. A gradual increase in incidence is observed in both groups; however, the Evolocumab group consistently exhibits a lower cumulative incidence compared to the Placebo group, indicating a potential protective effect. The numbers at risk displayed beneath the time axis represent the count of patients still under observation and stroke-free at each time point, thereby providing important context for interpreting the event probabilities over time.

A comparison between the reconstructed IPD-based cumulative incidence curves and those reported in the clinical study report (CSR), shown in Figure 2.3, reveals strong visual agreement. The shapes of the curves are closely aligned, supporting the validity of the data reconstruction process.

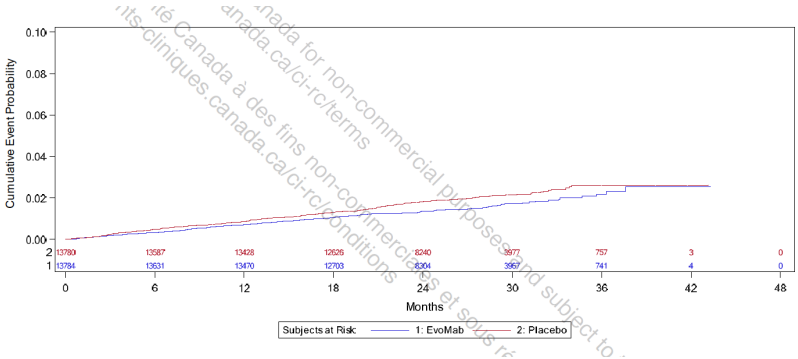

Figure 2.3: Cumulative incidence curves of stroke from CSR

While Table 2.4 and Figure 2.2 specifically illustrate the results for stroke, the same reconstruction procedure was applied to the remaining four cardiovascular event types, making up the composite endpoints: myocardial infarction, coronary revascularization, cardiovascular death, and hospitalization for unstable angina. Each reconstructed dataset follows the same structure and format, enabling consistent downstream analyses.

2.2 Sequential Events Simulation

In the previous section on data reconstruction, we obtained reconstructed survival life tables for each type of event, respectively. However, these summaries represent marginal survival experiences and do not capture the sequence of events experienced by individual patients. In this section, we focus on the reverse engineering process to simulate plausible individual-level trajectories by generating a series of sequential events for each patient, based on the reconstructed survival life tables.

2.2.1 Multi-State Model Framework

To better understand the mechanism underlying sequential events in disease progression, we adopt the concept of a multi-state model.

A multi-state model is an extension of traditional survival analysis that enables individuals to transition through multiple states over time, rather than focusing solely on the time to a single event. This framework provides a more realistic depiction of clinical progression, particularly in chronic or complex diseases where patients may experience several types of events before reaching a terminal state. By capturing transitions at the individual level, multi-state models allow for a richer representation of patient event histories.

Consider the FOURIER trial, which investigates cardiovascular outcomes using a composite endpoint composed of five types of clinical events: cardiovascular death, stroke, hospitalization for unstable angina, myocardial infarction, and coronary revascularization. In this context, as long as a patient has not yet experienced the terminal event (i.e., cardiovascular death), they remain at risk of developing one or more of the terminal or non-terminal events. Because of the relatively young patient population, we ignore competing death of other causes other than cardiovascular causes. The possible event trajectories for patients can be visualized using the multi-state diagram in Figure 2.4.

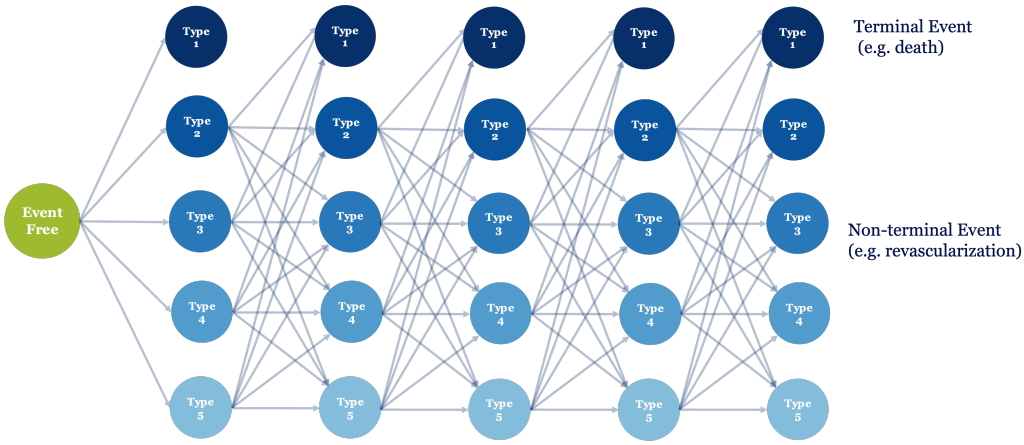

Figure 2.4: Multi-state Framework

In the specific context of the FOURIER trial, all patients start in the event-free state, depicted as the leftmost circle in Figure 2.4. The columns from left to right represent the sequence of transitions over time, illustrating how patients move from one clinical state to another. State 1 corresponds to the terminal event (cardiovascular death), while States 2 through 5 represent the various non-terminal events such as stroke, myocardial infarction, and other cardiovascular complications. The arrows between states indicate the possible transitions a patient can experience during the follow-up period. This structure enables modeling of complex clinical pathways and helps quantify the transition probability and timing of each event type.

2.2.2 Reverse Engineering Algorithm

This algorithm is designed to map the reconstructed type-specific life tables onto a multi-state structured life table, thereby enabling the simulation of individual patient trajectories. The core idea relies on the use of estimated type-specific hazard rates, augmented with an adjustment mechanism that compensates for event dependence. This adjustment allows us to simulate, for each patient, a plausible sequence of event times along with their corresponding event types, consistent with the underlying multi-state process.

Assumptions The reverse engineering algorithm is implemented under the following modeling assumptions, and these assumptions are made for both arms independently:

1. **Limited number of events ($e \leq 5$):** Each individual experience a maximum of five events during the follow-up period. This constraint reflects a practical upper bound informed by the clinical context and the duration of the trial.

2. **Constant type-specific hazard rates (h_k):** Each event type is associated with a fixed baseline hazard rate that remains constant over time.

3. **Increased risk after the first event ($c_k^1 < c_k^{2+}$; $c_k^1 < 1$)** This assumption states that once a patient experiences a non-terminal event, the risk of subsequent events of the same type increases. The risk is modeled through type-specific multiplicative adjustments to the baseline hazard rates, represented by compensation parameters c_k^e, where k denotes the event type and e the event order. Specifically, $c_k^1 < c_k^{2+}$ indicates that the first occurrence of an event carries a lower risk than subsequent occurrences, and $c_k^1 < 1$ indicates that the first-event risk is below the baseline risk accounting for the full set of five possible events.

4. **Fixed post-event acceleration ($c_k^2 = c_k^3 = c_k^4 = c_k^5 = c_k^{2+}$; $c_k^{2+} > 1$):** After the first non-terminal event, the hazard rates for all subsequent events are increased by a constant, event-type-specific multiplier, i.e., compensation parameter c_k^{2+}, which, remains fixed for the rest of the follow-up period, so that $c_k^2 = c_k^3 = c_k^4 = c_k^5 = c_k^{2+}$. And the risk of the subsequent events is higher than the baseline risk accounting for the full set of five possible events, so that $c_k^{2+} > 1$.

These assumptions help simplify the modeling process while still capturing key clinical dynamics, such as the increased risk following earlier complications. Moreover, they align with clinical insights and are considered realistic based on expert input.

Step 1. Estimate type-specific hazard rates In this step, we estimate the hazard rate for each event type separately within each treatment arm using reconstructed individual patient data (IPD). The hazard rate for event type is estimated as the ratio of the total number of observed events of a specific type to the total time at risk accumulated across all individuals in the corresponding treatment group.

$$\hat{h}_k = \frac{\#\text{Event}_k}{\sum_i \text{time}_i}, \tag{2.1}$$

where k represents the specific type of event, i represents the id of the patient. This computation is performed separately for each treatment arm (e.g., placebo and Evolocumab) and for each type of event (e.g., stroke, myocardial infarction, etc.).

Taking stroke (Figure 2.5) and revascularization (Figure 2.6) as examples, the dashed lines represent the corresponding estimated constant hazards, which align closely with the survival curves. This good agreement indicates that the constant hazard assumption is reasonable in these cases.

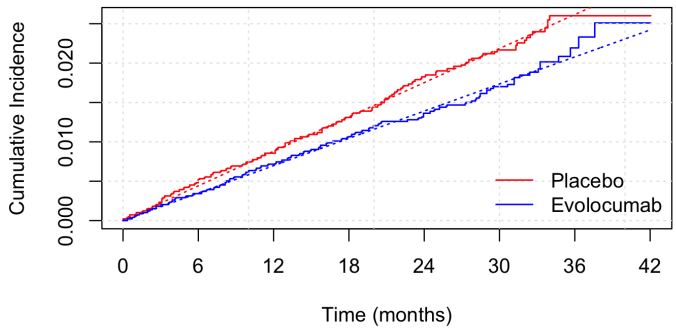

Figure 2.5: Estimated constant hazard of event Stroke

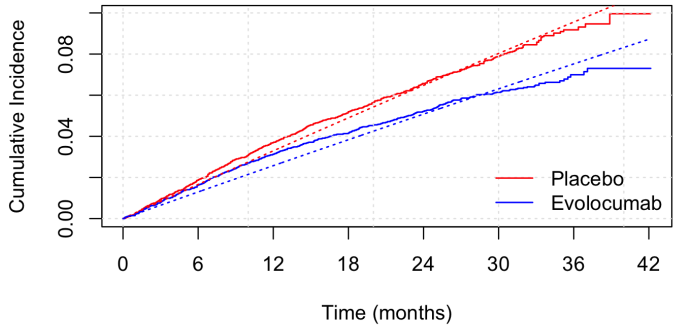

Figure 2.6: Estimated constant hazard of event Revascularization

Step 2. Generate a Multi-state Process by using the type-specific hazard as a reference.

To simulate a realistic multi-state process, we build on the previously estimated constant hazard rates h_k for each event type k across treatment arms. These baseline hazards represent the overall risk across the full follow-up period but do not reflect time-dependent changes in event susceptibility after prior occurrences. To address this, we introduce compensation factors—scalars applied to the baseline hazard to simulate the increased or decreased risk after previous events. Specifically, we define c_k^1 as the type-specific compensation for the first event, and $c_k^{2^+}$ as the compensation applied to subsequent events (second and beyond), and each treatment arm has its own independent compensation parameters. Accordingly, we assign initial compensation values such that $c_k^1 < 1$ (lower risk for first events) and $c_k^{2^+} > 1$ (higher risk for later events), which reflects the assumption of "Increased risk after the first event", that once a patient has experienced an event, they become more vulnerable and are more likely to experience additional events.

The generating mechanism works as follows: we adjust the baseline hazard by applying the compensation factor for the current event layer. For each individual, and for each event type, we sample a time to event from an exponential distribution with the adjusted hazard. The actual event is then determined by selecting the type corresponding to the shortest sampled time. After recording the first event, individuals who experience terminal events are removed from the risk set, while the remaining individuals proceed to the next event layer, where the same procedure is repeated. And the logic of this event generation process is implemented in Algorithm 1.

Algorithm 1 Simulation of Multi-State Process

Input: Number of individuals n, maximum follow-up time τ;
Hazard rates $h_1, h_2, \ldots, h_k$;
Compensation factors for first events c_k^1, and later events c_k^{2+}.
Output: Simulated life history data: `df` (uncensored), `df_c` (censored).

Initialize empty matrix `df`
for each individual $i = 1$ to n **do**
  Initialize time $t \leftarrow 0$, and event count $e \leftarrow 1$
  while $t < \tau$ and $e \leq 5$ **do**
    Set compensation factor $c_k \leftarrow \begin{cases} c_k^1 & \text{if } e = 1 \\ c_k^{2+} & \text{if } e > 1 \end{cases}$
    Sample event times: $t_k \sim \text{Exp}(\lambda_k), \quad \lambda_k = c_k \cdot h_k$
    Determine next event time: $t_{\text{next}} \leftarrow \min_k t_k$
    Identify event type: $d_{\text{next}} \leftarrow \arg\min_k t_k$
    Update total time: $t \leftarrow t + t_{\text{next}}$
    if $t \geq \tau$ **then**
      break
    end if
    Record (t, d_{next}) in `df` for individual n
    if $d_{\text{next}} =$ absorbing state **then**
      break
    end if
    Increment event count: $e \leftarrow e + 1$
  end while
end for
Sample censoring times: $t_c^{(i)} \sim \text{Uniform}(17, 36)$ for all i
Create `df_c` by censoring events in `df` at $t_c^{(i)}$
return `df` and `df_c`

To make the simulated data more reflective of real-world conditions, censoring was incorporated into the simulation framework. Based on the empirical pattern observed in the reconstructed individual patient data (IPD), censoring times appeared to follow an approximately uniform distribution. Accordingly, censoring times were modeled using a uniform distribution, specifically $U(17, 36)$, which provided a close approximation to the reconstructed IPD censoring behavior.

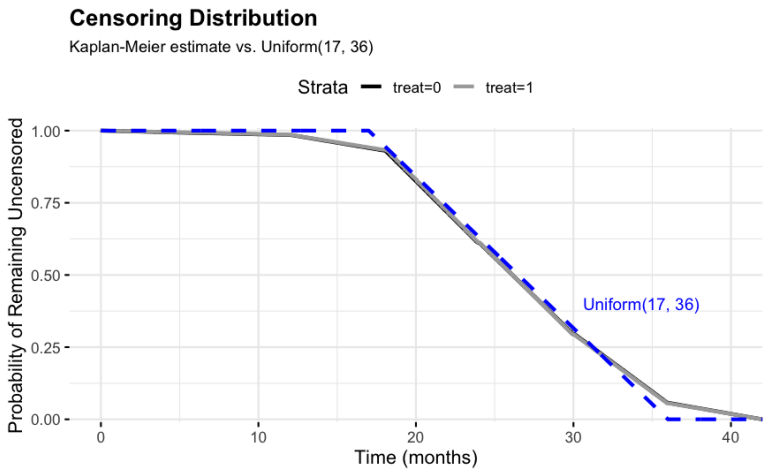

Figure 2.7: Find the censoring distribution

As shown in Figure 2.7, the estimated censoring distributions for the placebo and evolocumab groups (black and grey lines, respectively) are nearly identical and both closely match the proposed uniform distribution (blue dashed line). This close visual agreement supports the use of $U(17, 36)$ as a reasonable approximation for censoring in the simulated dataset. Furthermore, the median of $U(17, 36)$ is 26.5 months, which is consistent with the median follow-up time reported in the FOURIER trial (26.4 months, or 2.2 years).

At the end of this step, two datasets are returned: one containing the complete simulated multi-state event histories without censoring, and the other reflecting observed data after applying censoring.

Step 3. Tuning Compensation Parameters by minimizing the loss function.

In the earlier steps, we initialized the compensation parameters $\mathbf{c} = (c_k^1, c_k^{2+})$ to ensure the multi-state simulation could proceed. These initial values were based on rough estimates and were not yet adjusted to match the statistical summary from the original clinical data.

At this step, we improve the accuracy of the simulation by fine-tuning these parameters using an optimization method. The aim is to find a set of compensation parameters $\hat{\mathbf{c}}$ such that the simulated data reproduces several key metrics from the clinical report. These target metrics include:

- The number of first events of each type (revascularization, MI, hospitalization, stroke, Cardiovascular death) for each treatment arm;
- The total event counts across the entire follow-up period for each event type for each treatment arm, and;
- The distribution of event frequencies per patient, i.e., how many patients experienced exactly 1, exactly 2, or at least 3 events for each treatment arm.

To keep the parameter space manageable and reduce overfitting, we assume that the hazard adjustment (compensation) remains the same for all events beyond the second. That is, for each event type k, we set: $c_k^2 = c_k^3 = c_k^4 = c_k^5 = c_k^{2^+}$.

This leaves us with two parameters per event type: c_k^2 for the first event and $c_k^{2^+}$ for all subsequent events. In total, we optimize $2 \times 5 = 10$ compensation parameters.

The tuning procedure follows a simulation-based optimization approach. For a given set of parameters, the event generation algorithm (Algorithm 1) simulates event histories for a synthetic population. From the censored dataset produced, we extract summary statistics that correspond to our predefined targets. A loss function is then computed as the sum of absolute differences between the simulated values and the target values for each metric. This provides a scalar objective function that quantifies the mismatch of the statistical summaries between simulated and original data.

To find the optimal set of parameters, we use the Nelder-Mead method—a numerical optimization technique commonly used for nonlinear problems in multidimensional spaces. In our case, the loss function is nonlinear and depends on multiple factors, making Nelder-Mead an appropriate choice for efficiently navigating the parameter space to find a minimum. The entire tuning procedure is summarized in Algorithm 2.

Algorithm 2 Tuning of Compensation Parameters

Input: Initial compensation matrix $\mathbf{C} = \left(c_k^1, c_k^{2+}\right)_{5\times 2}$;
Estimated hazard vector $\hat{\mathbf{h}}_\mathbf{k} = (\hat{h}_1, \ldots, \hat{h}_5)$;
Target summary statistics:
First event counts $T^{(1)}$, total event counts $T^{(\text{total})}$, and subjects counts $T^{(\text{sub})}$.
Output: Optimized compensation matrix $\hat{\mathbf{C}}$

function Loss($\mathbf{C}, \hat{\mathbf{h}}_\mathbf{k}, T^{(1)}, T^{(\text{total})}, T_{(\text{sub})}$)
  Simulate multi-state data using **Algorithm 1** with inputs $\mathbf{C}, \hat{\mathbf{h}}_\mathbf{k}$
  Extract from censored dataset (`df_c`): ▷ Simulation-based estimates
    $\hat{T}_k^{(1)}$: first event type counts
    $\hat{T}_k^{(\text{total})}$: total event type counts
    $\hat{T}_{(sub)}^{j}$: subject counts by event count ($j = 1, 2, \geq 3$)
  Compute loss:

$$\mathcal{L} = \sum_{k=1}^{5} \left| \hat{T}_k^{(1)} - T_k^{(1)} \right| + \sum_{k=1}^{5} \left| \hat{T}_k^{(\text{total})} - T_k^{(\text{total})} \right| + \sum_{j=1}^{\geq 3} \left| \hat{T}_{(sub)}^{j} - T_{(sub)}^{j} \right| \tag{2.2}$$

  return $\mathcal{L}$
end function
function Optimize($\mathbf{C}_{\text{init}}, \hat{\mathbf{h}}_\mathbf{k}, T^{(1)}, T^{(\text{total})}, T_{(\text{sub})}$)
  Use Nelder–Mead simplex method to minimize Loss($\mathbf{C}, \hat{\mathbf{h}}_\mathbf{k}, \ldots$)
  return optimized matrix $\hat{\mathbf{C}}$
end function

Once the optimization converges, the best-fitting parameters $\hat{\mathbf{c}}$ for both arms are returned. These can then be used to generate simulated event histories that closely align with the target metrics.

Step 4. Quality Check by comparing Cox regression results from simulated data with published results.

To assess the quality of the simulated dataset, we perform a validation step by comparing the results of a Cox proportional hazards model applied to the simulated data against the results reported in the original clinical study.

In the published paper, the time-to-event analysis was based on the first occurrence of any composite endpoint, such as cardiovascular death, myocardial infarction, stroke, hospitalization, or revascularization. The cumulative incidence function used in the FOURIER study was based on the formula $1-S(t)$, where $S(t)$ is the Kaplan-Meier survival estimate. In our quality check, we follow the same methodology and do not account for competing risks, thereby staying consistent with the ignorance of the original analysis.

Figure 2.8 shows the (incorrect, competing risk was ignored) cumulative incidence curves from the simulated data, stratified by treatment group. The red line represents the placebo group and the blue line represents the treatment group (Evolocumab). The gap between the curves reflects a lower cumulative incidence of composite events in the treatment group over time, which is qualitatively consistent with the published findings.

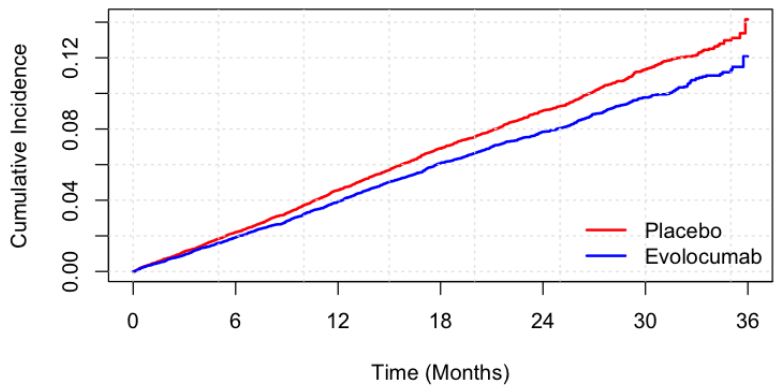

Figure 2.8: Cumulative incidence curves of composite endpoints with simulated data

To quantify this comparison, we fit a Cox model using the simulated data with the treatment arm as the covariate. The resulting hazard ratio (HR), confidence interval (CI), and p-value are shown in Table 2.5, alongside the corresponding values reported in the published study.

Table 2.5: Comparison of Hazard Ratios (HR)

Source	**HR**	**95% CI**	**P-value**
Uncensored Simulated Data	0.86	(0.81, 0.93)	< 0.001
Censored Simulated Data	0.86	(0.80, 0.93)	< 0.001
Published Paper	0.85	(0.79, 0.92)	< 0.001

The hazard ratio from the simulated censored data is 0.86, which is very close to the published value of 0.85. This represents only a 1% relative difference, with both results are statistically significant, and the confidence intervals of both estimates overlap substantially. Additionally, the Hazard Ratio result of censored and not censored data doesn't differ a lot, 0.8649 for the uncensored and 0.8630 for the censored to be more precise, which means the censoring doesn't have big impact on the Cox regression result in this case. These findings suggest that the simulation closely replicates the survival behavior observed in the actual clinical trial, indicating good quality and reliability of the generated data.

2.3 Result of Data Simulation

This section presents the outcomes of the simulation procedure introduced before. The goal of this simulation is to generate a synthetic dataset that preserves key statistical properties of the original trial, including the distribution of first and subsequent cardiovascular events, and the overall frequency of event counts per subject.

The following paragraphs present comparisons between the simulated data and the clinical report, and introduce a sample life table extracted from the synthetic dataset to illustrate multi-state event trajectories across representative patients.

Estimated Hazard Rates were calculated separately for each treatment group (Evolocumab and Placebo) and for each clinical event type considered in the trial. These estimated rates were derived from Equation (2.1), based on reconstructed IPD tables. The outcomes of estimated hazard rates listed in Table 2.6, which are expressed per 1,000 person-months.

Table 2.6: Estimated Hazard Rate (per 1000 person-months)

Event Type	**Evolocumab**	**Placebo**	**Ratio**
Revascularization	2.172	2.790	0.779
Myocardial Infarction	1.333	1.828	0.729
Hospitalized Angina	0.538	0.557	0.966
Stroke	0.583	0.737	0.791
Cardiovascular Death	0.682	0.664	1.027

Note that the hazard ratio for cardiovascular death slightly exceeds 1, indicating a marginally higher event rate in the Evolocumab group compared to placebo. In contrast, the hazard ratios for all other event types: myocardial infarction (MI), revascularization, stroke, and hospitalized angina are below 1, suggesting an underlying protective effect of Evolocumab in reducing the risk of these non-fatal cardiovascular events. However, not all components of the composite endpoint are benefitted equally.

Optimized Compensation Parameters As described before, compensation parameters were introduced to adjust the event rates in the simulated multi-state process to better match the clinical summaries. As we mentioned before, these parameters were optimized separately for the placebo and evolocumab groups by minimizing a loss function that incorporates event type distributions and the number of patients experiencing sequential events.

Table 2.7 below presents the final optimized values of the compensation parameters. For each clinical event type, we report the estimated compensation for the first event c_k^1 and for all subsequent events c_k^{2+}.

Table 2.7: Optimized Compensation Parameters

Event Type k	**Placebo**		**Evolocumab**	
	$\boldsymbol{c_k^1}$	$\boldsymbol{c_k^{2+}}$	$\boldsymbol{c_k^1}$	$\boldsymbol{c_k^{2+}}$
Revascularization	0.38	13.42	0.44	15.19
Myocardial Infarction	0.63	7.33	0.67	7.35
Hospitalized Angina	0.89	9.04	0.95	12.43
Stroke	0.83	4.25	0.85	3.47
Cardiovascular Death	0.67	10.11	0.66	12.23

These parameters serve as adjustments to the underlying hazard structure and enable the simulation to more closely mimic the reported clinical data distributions. The variation in c_k^1 and c_k^{2+} between treatment arms reflects differences in event dynamics due to treatment effect.

Table 2.8: Number of First and Total Events of Simulated Data

Event Type	**Placebo**		**Evolocumab**	
	First	**Total**	**First**	**Total**
Revascularization	386	965	353	769
Myocardial Infarction	418	629	325	444
Hospitalized Angina	167	233	168	238
Stroke	225	274	191	220
Cardiovascular Death	152	241	153	254
Total	**1,348**	**2,342**	**1,190**	**1,925**

Table 2.9: Number of First and Total Events in Clinical Report

Event Type	**Placebo**		**Evolocumab**	
	First	**Total**	**First**	**Total**
Revascularization	394	965	349	759
Myocardial Infarction	423	639	329	468
Hospitalized Angina	160	239	169	236
Stroke	226	262	184	207
Cardiovascular Death	142	240	161	251
Total	**1,345**	**2,345**	**1,192**	**1,921**

Under the optimized set of compensation parameters, the resulting simulated event counts are summarized in Table 2.8. This table presents the number of first and subsequent events by clinical event type for both the placebo and evolocumab groups. To investigate how well the simulation reflects the actual trial outcomes, these simulated counts are directly compared with the event counts extracted from the clinical study report, shown in Table 2.9.

A close inspection of the two tables reveals that the differences between simulated and original ones are generally small. For example, in the placebo group, the number of first events for myocardial infarction is 418 in the simulation versus 423 in the clinical report, and the subsequent event counts are nearly identical across most event types. Similarly, in the evolocumab group, the simulated first event count for revascularization is exactly matched to the clinical report (349 events), and other categories show only minor deviations.

The total number of events—both first and subsequent—also aligns very closely with the reported figures. For instance, the placebo group shows a total of 1,348 first events and 2,342 total events in the simulation, compared to 1,345 and 2,345 in the clinical report. Such small differences (3 events or fewer in total counts) are negligible in the context of over 13,000 patients per group and suggest that the optimized simulation effectively captures the structure and intensity of event occurrences.

Let's turn our attention to another key dimension of optimization target: The distribution of event frequencies.

Table 2.10: Comparison of Event Frequency Distributions

Group	Event Count	Simulated	CSR
Placebo (n=13780)	1 Event	704	806
	2 Events	399	543
	$\geq$3 Events	245	214
	Total	**1348**	**1563**
Evolocumab (n=13784)	1 Event	677	768
	2 Events	348	417
	$\geq$3 Events	165	159
	Total	**1190**	**1344**

Table 2.10 further shows the performance of the optimized simulation by comparing the distribution of the number of events per patient with those reported in the clinical study. The comparison shows a close alignment between the simulated and reported data, suggesting that the optimization procedure successfully captured the real event structure.

Specifically, the differences between simulated and clinical data across the categories of patients experiencing 1, 2, or $\geq$3 events are relatively minor. For both treatment arms, the total number of patients with events is slightly lower in the simulation than in the clinical report, but the proportions remain consistent. For example, in the placebo group (n = 13,780), approximately 10.0% of patients had at least one event in the simulated data (1,348 / 13,780), while in the evolocumab group (n = 13,784), the proportion was 8.6% (1,190 / 13,784). These values are closely aligned with the proportions in the clinical report (11.3% and 9.8%, respectively).

These results also highlight that over 90% of patients did not experience any events during the follow-up period, which aligns with clinical expectations in rare-event cardiovascular outcomes.

Recall that the optimization of the compensation parameters was driven by a custom-defined loss function:

$$\mathcal{L} = \sum_k \left| \hat{T}_k^{(1)} - T_k^{(1)} \right| + \sum_k \left| \hat{T}_k^{(\text{total})} - T_k^{(\text{total})} \right| + \sum_j \left| \hat{T}_{(\text{sub})}^{j} - T_{(\text{sub})}^{j} \right|$$

This loss function penalizes deviations between the simulated and observed counts of first events, total events by type, and distributions of event frequencies. Under this framework, the resulting total loss was 337 for the placebo group and 235 for the evolocumab group.

To contextualize these values, it is helpful to consider the limitations in the clinical report itself. Specifically, if a patient experienced multiple events on the same day, the report did not specify which event occurred first. According to the trial documentation, 367 such ambiguous same-day events were observed in the placebo group and 271 in the evolocumab group. These ambiguities impose an upper bound on the achievable accuracy of any simulation that attempts to match detailed event timing. Given that the calculated losses (337 and 235) are both within these bounds, they are considered reasonable and acceptable.

Simulated Sequential Event Life table Using optimized compensation parameters and constant hazard rates, we simulated multi-state recurrent event data under two treatment arms: placebo and evolocumab. The resulting dataset consists of longitudinal patient records, with each record capturing up to five sequential events, their corresponding event types, and the times at which these events occurred.

It is important to note that the majority of patients did not experience any events during the follow-up period. To illustrate more informative event trajectories, a subset of selected patient records is presented in Table 2.11. These rows show sequential transitions through different event states over time.

Table 2.11: Sequential Event Survial Life Table of Placebo Group(Time in Months)

ID	Time1	Type1	Time2	Type2	Time3	Type3	Time4	Type4	Time5	Type5
1	36.00	0	–	–	–	–	–	–	–	–
4	19.86	5	–	–	–	–	–	–	–	–
9	6.72	3	6.72	1	24.84	2	27.68	5	–	–
14	21.99	2	23.35	5	–	–	–	–	–	–
16	5.94	1	19.40	1	33.29	0	–	–	–	–
23	2.91	4	13.12	1	29.76	5	–	–	–	–
528	2.73	4	9.08	3	12.88	3	23.21	3	30.12	5
616	5.48	4	6.99	3	8.60	3	13.92	3	23.53	1

Note 1: Subject of ID 1 is an example of no events until full follow-up time,
Note 3: Type1 means the type of event occurs for Time1, and so on,
Note 3: 0 = Censored, 1 = Revascularization, 2 = Myocardial Infarction , 3 = Hospitalized Angina, 4 = Stroke, 5 = Cardiovascular Death, , for the event type coding

Although the individual patient trajectories from the real study remain inaccessible, this simulation reconstructs one plausible realization of the underlying data-generating process, guided by aggregate-level statistics reported in the clinical trial. The synthetic data not only mirrors the empirical distributions of event counts and types but also enables further downstream analysis, including survival modeling, quality-adjusted life year (QALY) calculations, and treatment effect estimation.

Another additional step to the Data Simulation, in order to fit the framework of package `mstate` [9], we defined the function to transfer long format data to wide format one, the function will be included in the appendix. To show the result in long format transferred from the simulated data in wide format in Table 2.11, to make the transfer more understandable, we take example of patient id 23 to illustrate. After the format transforming function, the wide format table will be transfered into long format which shows in the following Table 2.12:

start	stop	status	from	to	transition	time	transno	transname
0.00	2.91	0	1	1	1→1	2.91	–	–
0.00	2.91	0	1	2	1→2	2.91	1	Event-free → Revasc
0.00	2.91	0	1	3	1→3	2.91	2	Event-free → MI
0.00	2.91	0	1	4	1→4	2.91	3	Event-free → Hosp
0.00	2.91	1	1	5	1→5	2.91	4	Event-free → Stroke
0.00	2.91	0	1	6	1→6	2.91	5	Event-free → Death
2.91	13.12	0	5	1	5→1	13.12	–	–
2.91	13.12	1	5	2	5→2	13.12	18	Stroke → Revasc
2.91	13.12	0	5	3	5→3	13.12	19	Stroke → MI
2.91	13.12	0	5	4	5→4	13.12	20	Stroke → Hosp
2.91	13.12	0	5	5	5→5	13.12	–	–
2.91	13.12	0	5	6	5→6	13.12	21	Stroke → Death
13.12	29.76	0	2	1	2→1	29.76	–	–
13.12	29.76	0	2	2	2→2	29.76	–	–
13.12	29.76	0	2	3	2→3	29.76	6	Revasc → MI
13.12	29.76	0	2	4	2→4	29.76	7	Revasc → Hosp
13.12	29.76	0	2	5	2→5	29.76	8	Revasc → Stroke
13.12	29.76	1	2	6	2→6	29.76	9	Revasc → Death

Note 1: 1 = event-free, 2 = Revascularization, 3 = Myocardiol Infarction , 4 = Hosptalized Angina, 5 = Stroke, 6 = Cardiovascular Death, for the state coding
Note 2: Death represents cardiovascular death, Hosp represents hospitalized Angina, and Revasc represents Revascularization

Table 2.12: Long format Life table of subject ID 23

Here the `transno` represents transition number, i.e., coded the possible transitions in Table 3.2 from numebr 1 (i.e., `state 1` → `state 2`) to number 21 (i.e., `state 5` → `state 6`).

Chapter 3

Q-RMST Method

In this chapter, we introduce a novel concept called the Q-RMST, short for QALY-weighted Restricted Mean Survival Time. This method addresses the limitations of traditional risk-based approaches by providing a more interpretable and comparable estimate, expressed in terms of patient life-years. Furthermore, it incorporates subsequent events and accounts for the varying clinical importance of different event types, resulting in a more precise and objective evaluation of treatment effects.

3.1 QALY Accumulated Weight Function

In this section, a core definition of the QALY accumulated weight function will be given, which adapts a health economics-related concept: Quality-Adjusted Life Year (QALY). By incorporating QALY into the weight function, the model can take the different importance of clinical outcomes into account.

3.1.1 Quality-Adjusted Life Year (QALY)

Quality-Adjusted Life Year (QALY) is a widely used metric in health economics for assessing the value of medical interventions. It combines both the quantity and quality of life into a single measure, where one QALY corresponds to one year of life lived in perfect health. This measure allows for the comparison of treatments not only based on survival outcomes but also on their impact on patients' quality of life.

QALY untility Each health state in the multi-state model is assigned a utility weight ranging from 0 to 1, where 0 represents death and 1 corresponds to perfect health. These weights reflect the relative desirability or quality of each state, typically derived from patient-reported outcomes, expert assessments, or population-based valuation studies.

Let $u_h \in [0, 1]$ denote the utility weight assigned to health state h. At any given time t, a patient's QALY contribution is determined by the state they occupy and its corresponding utility. For the FOURIER trial, we applied the utility values for each clinical state as described by Wouters et al. [10]. To visually compare the relative quality-of-life impacts across states, we present a bar plot in Figure 3.1:

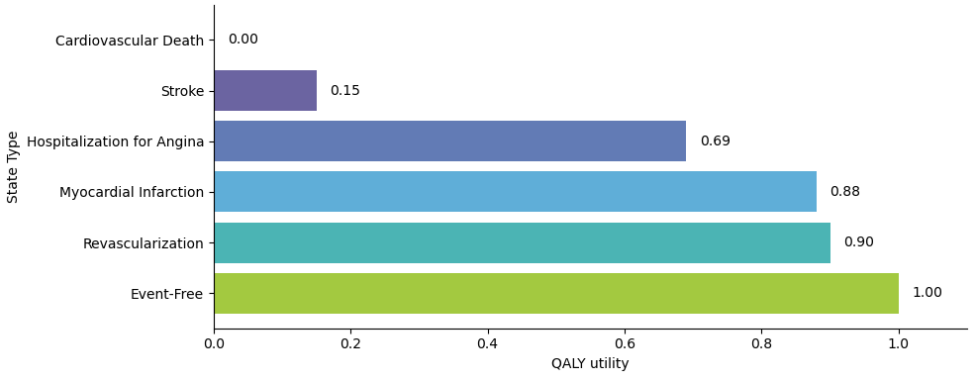

Figure 3.1: Barplot of QALY utilities

It is important to note that the utility value assigned to the stroke state represents the heaviest one among three subtypes: ischemic stroke, hemorrhagic stroke, and transient ischemic attack (TIA). The range of utility values associated with these subtypes will be further explored in the sensitivity analysis section.

3.1.2 QALY Accumulation Principles

To incorporate QALYs into survival analysis, particularly in the context of possible multiple events occurring in sequence, it is essential to define how QALY utility values are accumulated over time. Different accumulation principles reflect different assumptions about how patients experience and perceive health-related quality of life.

To formalize this, we define the QALY accumulated weight function, denoted as $w_Q(t)$, which specifies how a quality-adjusted value is assigned to each point in time based on a patient's health history. This weight function plays a central role in computing quality-adjusted survival metrics. Four principles were incorporated for QALYs accumulation, and we assumed that QALYs remain stable over time.

1. Markov Principle This principle assumes that the QALY contribution at each time point depends only on the patient's current health state. It reflects a memoryless process, where past states do not influence present utility. This is the most commonly applied in state-transition models. And under this principle the QALY accumulated weight function could be written as:

$$w_Q^1(t) = u_{X(t)},$$

where $X(t)$ is the number of state occupied at time t.

2. Worst-State Update Principle This principle assumes that the patient's perceived quality of life at each time point is determined by the most severe health state experienced up to that time. It captures the idea that serious events (e.g., stroke or myocardial infarction) may have a lasting negative impact on perceived health. Under this principle the QALY accumulated weight function could be written as:

$$w_Q^2(t) = \min_{0 \le s \le t} \{u_{X(s)}\}$$

3. Cumulative Product Principle This principle assumes compounding effect of sequential health states on overall quality of life. QALY is calculated as the product of utility weights across all experienced states. Under this principle the QALY accumulated weight function could be written as:

$$w_Q^3(t) = \prod_{s=0}^{t} u_{X(s)}$$

4. Conventional RMST Principle This is the conventional RMST approach, where QALY utilities are not incorporated at all. It treats all time spent alive equally, regardless of health state. Under this principle the QALY accumulated weight function could be written in a piecewise function form:

$$w_Q^4(t) = \begin{cases} 1, & \text{if } u_{X(t)} = 1 \\ 0, & \text{if } u_{X(t)} < 1 \end{cases}$$

3.2 Q-RMST Model Specification

In this section, we present the model specification of the Q-RMST approach, which integrates the QALY-accumulated weight function into the conventional Restricted Mean Survival Time (RMST) framework. We consider two scenarios: the first represents a more ideal situation, where patients are followed up without any censoring, relying on complete follow-up data. The second scenario reflects a more realistic setting, where censoring occurs. To address the challenges introduced by censoring, we again adopt the Multi-State Model framework.

3.2.1 Typical Example for the No-Censoring Scenario

To illustrate how different QALY accumulation principles influence the calculation of quality-adjusted survival time, we present a simplified example involving a single individual. For clarity, we assume complete follow-up over 36 months, with no censoring.

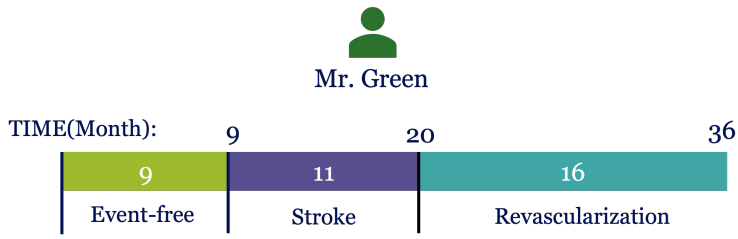

Figure 3.2: An example of the patient's sequential event trajectory over 36 months.

Figure 3.2 presents the clinical trajectory of a hypothetical patient who enters the trial in an event-free state. At month 9, the patient experiences a stroke, followed by a myocardial infarction (MI) at month 20. By the end of the 36-month follow-up, the individual had experienced two events and occupied a total of three distinct health states.

Table 3.1 summarizes the results of QALY-weighted survival times under four accumulation principles and outlines the key characteristics of each approach. It illustrates how different QALY accumulation principles can lead to markable different results, even when applied to the same health trajectory. In the toy example, the Markov principle yields the highest QALY (24.73 months), as it fully credits periods of partial recovery. In contrast, the Leave-Q-Out approach produces the lowest value (9 months), counting only time spent in perfect health and disregarding all other periods entirely. The Worst-State Update and Cumulative Product methods fall in between (13.05 months and 12.762 months, respectively), each introducing a form of memory or compounding to reflect the lasting or sequential impact of past health states. These differences highlight how the choice of weighting principle can significantly influence the interpretation of treatment benefit.

Method	Weight Function	Result	Strength	Limitation
Markov	$w_Q^1(t) = u_{X(t)}$	25.05	Simple and interpretable	Ignores past events
Worst-State Update	$w_Q^2(t) = \min_{0\le s\le t}\{u_{X(s)}\}$	13.05	Captures lasting negative effects	May overpenalize severe states
Cumulative Product	$w_Q^3(t) = \prod_{s=0}^{t} u_{X(s)}$	12.81	Models the trajectory sensitively	May undervalue recovery
Conventional RMST	$w_Q^4(t) = \begin{cases} 1, & u_{X(t)} = 1 \\ 0, & u_{X(t)} < 1 \end{cases}$	9	Straightforward and intuitive	Disregards all imperfect states

Calculation 1 (Markov): $1 \cdot 9 + 0.15 \cdot 11 + 0.9 \cdot 16 = 25.05$,
Calculation 2 (The Worst-update): $1 \cdot 9 + 0.15 \cdot 11 + 0.15 \cdot 16 = 13.05$,
Calculation 3 (Cumulative Product): $1 \cdot 9 + 0.15 \cdot 11 + 0.15 \cdot 0.9 \cdot 16 = 12.81$,
Calculation 4 (Conventional RMST): $1 \cdot 9 = 9$.

Table 3.1: Comparison of QALY accumulation principles

Each of these approaches represents a distinct way of interpreting the relationship between health state transitions and quality of life over time. The choice of principle depends on clinical context and modeling objectives.

3.2.2 Adapted Multi-State Model for the Censoring Scenario

In Section 2.2.1, we introduced the framework of the multi-state model to illustrate the mechanism underlying the simulation procedures. In this chapter, we continue to employ the concept of the multi-state model, but from a different analytical perspective.

Multi-state Model Specification

In clinical trials, it is common for an individual's complete event history to be only partially observed. Censoring presents a significant challenge in estimating treatment efficacy without bias. Multi-state model addressed the censoring challenge by modeling the transition hazards between states.

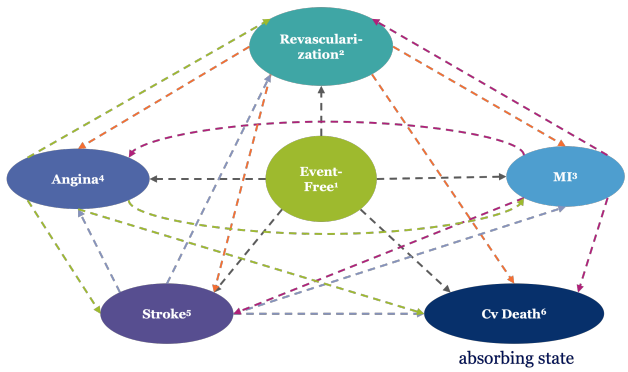

Figure 3.3: Multi-state Model

Figure 3.3 illustration of a Multi-State Transition Structure, specifically for the FOURIER Trial. Where arrows indicate the possible transitions between states.

We now formalize the multi-state model using three key components: transition hazards, transition probabilities, and expected length of stay (ELoS).

Let $X(t)$ be a stochastic process representing the state of an individual at time t, with the state space of $\{1, 2, ..., K\}$, denoted as $X(t) \in \{1, 2, ..., K\}$. In the FOURIER trial setting, patients may transition between states according to 21 possible transitions, summarized in Table 3.2. Self-transitions (remaining in the same state) are not explicitly listed in the matrix but are implicitly understood as the individual remaining in that state.

		To (state h)					
		Event Free	**Revasc.**	**MI**	**Hosp. Angina**	**Stroke**	**Cv Death**
From (state g)	**Event Free**	–	1→2	1→3	1→4	1→5	1→6
	Revasc.	–	–	2→3	2→4	2→5	2→6
	MI	–	3→2	–	3→4	3→5	3→6
	Hosp. Angina	–	4→2	4→3	–	4→5	4→6
	Stroke	–	5→2	5→3	5→4	–	5→6
	Cv Death	–	–	–	–	–	–

Table 3.2: Transition Matrix

Table 3.2 illustrates the Transition matrix, which specifying allowed transitions between clinical states in the FOURIER trial.

Transitions between states are governed by the transition hazard function $\lambda_{gh}(t)$, which quantifies the instantaneous risk of moving from state g to state h at time t, given that the individual is currently in state g. Formally:

$$\lambda_{gh}(t) = \lim_{dt \to 0} \frac{P(X(t + dt) = h) \mid X(t) = g)}{dt}, \tag{3.1}$$

where $X(t)$ is a time-inhomogeneous stochastic process mentioned before.

The transition probability $P_{gh}(s, t)$ is the probability that an individual is in state h at time t, given they were in state g at time s, which is defined as:

$$P_{gh}(s, t) = P(X(t) = h \mid X(s) = g) \tag{3.2}$$

It gives a time-dependent measure of how likely it is to go from one state to another.

Aalen–Johansen estimator

To estimate transition probabilities in the presence of censoring and competing risks, we use the Aalen–Johansen estimator—a matrix-based extension of the Kaplan–Meier estimator designed for multi-state models.

This estimator handles right-censoring by aggregating observed transitions and risk sets across time, which given the estimated transition hazard is:

$$\hat{\lambda}_{gh}(t) = \frac{dN_{gh}(t)}{Y_g(t)} \tag{3.3}$$

where,

- $Y_g(t)$: number of individuals at risk of transitioning from state g at time t,

- $N_{gh}(t)$: number of observed transitions from state g to state h by time t.

By integrating these estimates over time, we obtain the transition probability matrix:

$$\hat{P}(s,t) = \prod_{(s,t]} \Big(I + d\hat{A}(u) \Big) \tag{3.4}$$

where $\hat{A}(u)$ is the cumulative hazard matrix ($d\hat{A}_{gh}(u) = \hat{\lambda}_{gh}(t)$, and on the diagonal $d\hat{A}_{gg}(u) = -\sum_{h \neq g} d\hat{A}_{gh}(t)$), and I is the identity matrix.

This non-parametric method provides a robust estimate of how individuals move through different health states over time, accounting for censoring and competing events.

Expected Length of Stay (ELoS)

The Expected Length of Stay (ELoS) quantifies how long, on average, an individual will spend in a particular state h between times s and τ, assuming they were in state g at time s:

$$E_{gh}(s,\tau) = \int_s^\tau P_{gh}(s,t)\, dt \tag{3.5}$$

In practice, τ typically denotes the end of the follow-up period. ELoS helps researchers understand the impact of a disease and evaluate how treatments may affect the time spent in each state.

For the FOURIER trial, the formula for Expected Length of Stay (ELoS) can be simplified due to the fact that all patients begin in the same initial state: Event-Free. This common starting point allows us to drop the conditioning on the initial state and use a simplified form of the ELoS:

$$E_h(\tau) = \int_0^\tau P_h(t)\, dt \tag{3.6}$$

Here, $P_h(t) = P_{1h}(0,t)$ denotes the state occupation probability for state h, given that the process started in state 1 (the initial, event-free state) at time 0. This simplified notation will be used throughout the remainder of the analysis.

Q-RMST Model

To formulate the Q-RMST model, we take the product of the QALY-accumulated weight function and the Expected Length of Stay (ELoS) across different transitions:

$$\text{Q-RMST}(\tau) = \sum_{h=1}^{K} w_Q^i(\tau) \cdot E_h(\tau) \tag{3.7}$$

With τ denotes the time horizon of interest; $w_Q^i(\tau)$ is the QALY accumulated weight function of i principle; and $E_h(\tau)$ is the Expected Length of Stay of h state uo to time τ.

The total number of term K depends on the structure of the model and the chosen weighting principle. Under the Markov principle, the process is memoryless and only the current state influences utility. In this case, $K = 6$, the total number of states, as shown in the Multi-State Transition Diagram (Figure 2.4).

For the Worst-State Update and Cumulative Product principles, the memoryless assumption no longer holds. Instead, the full sequence of prior events must be taken into account to determine the appropriate weight at each time. This requires modeling the process as an extended multi-state model, where each distinct sequence of states (i.e., transitions over time) in the original multi-state model (Figure 2.4) constitutes a unique state in the extended multi-state model. Figure 3.4 visualizes this idea, with each state in the extended multi-state model representing a possible trajectory of state transitions.

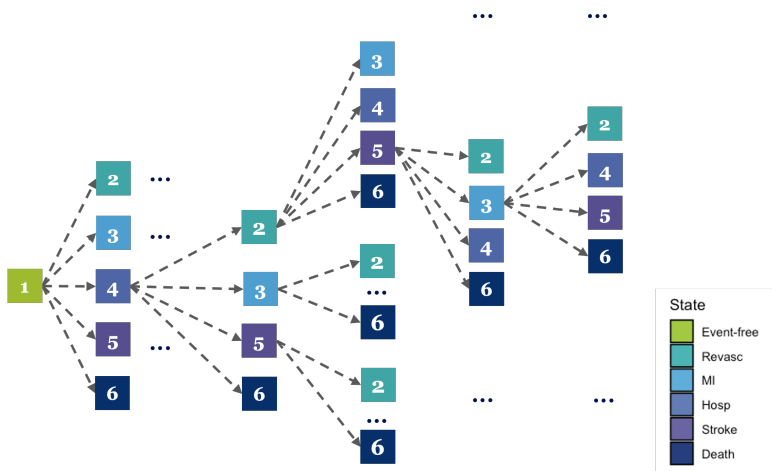

Figure 3.4: Extended Multi-state Model

In the context of the FOURIER trial, we define six clinically relevant states, including one absorbing state (cardiovascular death). Patients can experience up to five transitions during follow-up, meaning the maximum possible sequence length is five. In the extended multi-state model, K is not a single numeric index but a sequence of states. For example, $K = 12$ indicates a transition from state 1 to state 2, whereas $K = 13425$ represents a sequence of visits to states 1, 3, 4, 2, and 5 in that order.

3.3 Uncertainty Analysis

Uncertainty in our framework stems from the estimation of constant hazard rates used in the data simulation. As shown in Figure 2.5 and Figure 2.6, the estimated constant hazard lines align more closely with the survival curve for stroke than for revascularization. However, neither provides a perfect fit. This discrepancy highlights the inherent uncertainty in estimating constant hazard rates and suggests potential deviations from our underlying assumptions.

To capture the uncertainty, we introduce a perturbation term to the estimated hazard rate. Specifically, we draw the perturbation from a standard normal distribution $N(0, 1)$, scaled by the standard error of the log hazard estimate. First, recall how the constant hazard rate is estimated:

$$\hat{h}_k = \frac{\#\text{Event}_k}{\sum_i \text{time}_i}$$

We then define the perturbed hazard as:

$$\tilde{h}_k = exp(log(\hat{h}_k) + rnorm(1) \cdot se(log(\hat{h}_k))), \tag{3.8}$$

with $se(log(\hat{h}_k)) = \frac{1}{\#\text{Event}_k}$.

The uncertainty, carried by the perturbed hazard rates, is propagated throughout the data simulation process and the Q-RMST calculation, as illustrated in Figure 3.5.

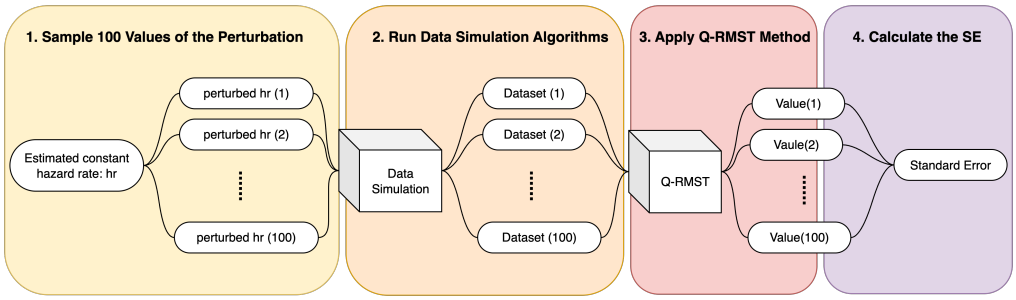

Figure 3.5: A pipeline to visualize the propagation of the perturbed hazard

1. Sample 100 values of the perturbation term $rnorm(1) \cdot se(log(\hat{h}_k))$ using different random seeds, resulting in 100 perturbed hazard vector $\tilde{h}_k$ for each arm.
2. Re-run data simulation algorithm for each perturbed hazard, including the optimization of compensation parameters. For each run, a new set of optimal compensation parameters was obtained. This step required substantial computing time due to the execution of 100 simulation iterations per perturbation.
3. Fit each new simulated dataset to the Q-RMST model, and yield 100 Q-RMST estimations for each treatment group.
4. Calculate the standard error of the estimated Q-RMST values.

The variability across these 100 replicates is then used to quantify the uncertainty of Q-RMST. Specifically, we compute the standard error (se) from the sampled Q-RMST values and construct 95% confidence intervals as:

$$95\%\ \text{CI} = (\widehat{\text{Q-RMST}} - 1.96 \cdot se,\ \widehat{\text{Q-RMST}} + 1.96 \cdot se)$$

This procedure allows us to formally account for uncertainty in the hazard rate estimation and propagate it through to our final Q-RMST results.

Chapter 4

Results

This chapter presents the results from the methods outlined in Chapter 3, covering the multi-state model outputs, Q-RMST estimates with associated uncertainty analyses, sensitivity analyses of QALY values, and a summary of findings from the FOURIER trial. The analyses are based on simulated, uncensored data with complete follow-up to 36 months, representing an idealized scenario that enables unbiased estimation of Q-RMST while avoiding complications introduced by censoring.

4.1 Multi-state Model: Expected Length of Stay

As introduced in Section 3.2.2, the Expected Length of Stay (ELoS) quantifies the average amount of time patients are expected to spend in each health state over a specified follow-up period. Mathematically, ELoS is computed as the integral of the state occupation probability over time, as shown in Equation (3.5).

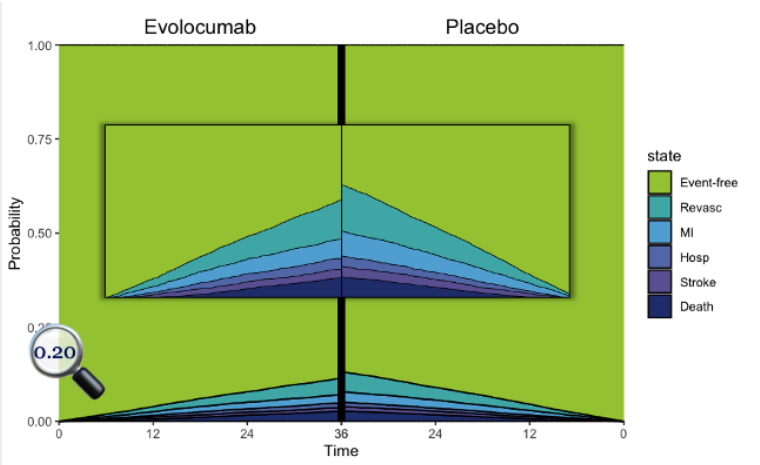

Figure 4.1: ELoS under Markov Principle

Figure 4.1 displays the estimated ELoS for both the Evolocumab and Placebo arms of the FOURIER trial, based on the Markov principle. To better visualize differences in rarely occupied health states, such as Stroke or Cardiovascular Death, a zoomed-in panel is included in the figure, focusing on the probability interval $[0, 0.2]$. This inset enables clearer comparison of expected durations in less common but clinically significant states.

While ELoS plots could, in principle, be generated under alternative weighting frameworks (e.g., Worst-State Update or Cumulative Product), these approaches involve a substantially larger and more complex set of state transitions. To maintain interpretability and focus, the EloS results are therefore discussed exclusively under the Markov principle.

Table 4.1 provides a numerical summary of the ELOS for a 36-month follow-up period, assuming all patients begin in the Event-Free state (state 1). For instance, the expected time spent in the Cardiovascular Death state in Evolocumab group is estimated to be 0.42 months. This does not imply that patients "remain alive while dead," but instead reflects an average across the full cohort: patients who do not die during follow-up contribute 0 months to this state, while those who die contribute $36 - t_6$ months, where t_6 is the time of death. For example, a patient who dies at month 35 contributes 1 month to the death state. Averaging these values across all individuals yields the final estimate.

Table 4.1: Expected Length of Stay (in Months) for patients in FOURIER Trail

To State	**Evolocumab**	**Placebo**
Event-Free (1)	33.89	33.57
Revascularization (2)	0.76	0.92
Myocardial Infarction (3)	0.47	0.58
Hospitalization for Angina (4)	0.24	0.23
Stroke (5)	0.22	0.26
Cardiovascular Death (6)	0.42	0.45
Total	36.00	36.00

From a broad perspective, the Event-Free state (see Table 4.1 and the yellow region in Figure 4.1) dominates the expected length of stay (ELOS) in both treatment arms, with patients in both groups spending approximately 34 months without experiencing any cardiovascular event. This finding indicates that the majority of patients remain event-free over the trial follow-up period.

Specifically, the Evolocumab group consistently shows reduced expected time spent in the following adverse health states compared to the placebo group: Revascularization, Myocardial Infarction, Stroke, and Cardiovascular Death state. The only exception is Hospitalization for Angina, where the expected durations are nearly identical between the two groups (0.24 vs. 0.23 months).

It is also worth noting that, while Table 2.8 reports a slightly higher number of cardiovascular deaths in the Evolocumab group compared to the Placebo group, the expected length of stay in the Cardiovascular Death state tells a different story. Specifically, patients in the Evolocumab group spend less time on average in this state (0.42 months vs. 0.45 months). This suggests that although cardiovascular deaths may occur more frequently in the Evolocumab group, they tend to happen later during follow-up. As a result, the time spent in the death state, which only begins after the event occurs, is shorter.

4.2 Result for Q-RMST estimation

To evaluate the treatment effect, we estimated the QALY-weighted Restricted Mean Survival Time (Q-RMST) using the reconstructed dataset without censoring. Since the data had no censoring, the adapted multi-state model was not required; instead, we applied the calculation procedure described in Section 3.2.1.

To assess the uncertainty of the Q-RMST estimates, we followed the method outlined in Section 3.3, which accounts for the variability in the estimated constant hazards used in the simulation. Confidence intervals for the treatment effect estimates were computed based on standard error propagation from the estimated Q-RMSTs in each arm.

Table 4.2 presents the estimated Q-RMST values for both the Evolocumab and Placebo groups under four different QALY-weighting principles: the Markov, the Worst-Update, the Cumulative Product, and conventional RMST approach. In each case, the treatment effect is shown as the difference in Q-RMST (Evolocumab minus Placebo), along with its associated standard error and 95% confidence interval.

Table 4.2: Q-RMST Results (in Months)

Principle	**Q-RMST**		**Difference**		
	Evolocumab	Placebo	Evo - Pla	se	95% CI
Markov	35.19	35.10	0.09	0.13	[-0.17, 0.34]
Worst Update	35.07	34.94	0.13	0.13	[-0.13, 0.39]
Cumulative Product	35.00	34.85	0.15	0.13	[-0.11, 0.42]
Conventional RMST	33.89	33.57	0.32	0.19	[-0.05, 0.68]

Overall, all four QALY weighting approaches suggest a modest advantage of Evolocumab over Placebo in terms of QALY-weighted restricted mean survival time. However, none of the observed differences reach statistical significance at the 5% level, as all confidence intervals include zero.

To compare across principles:

- The conventional RMST approach yields the largest estimated difference 0.32 months (9.56 days) and the widest confidence interval. This principle, which does not account for time-varying health states, may overestimate the benefit;

- The Markov, Worst-state Update, and Cumulative Product principles yield smaller effect sizes, ranging from 0.09 to 0.15 months (1.38 to 2.06 days), and consistent standard errors, reflecting a more conservative estimation of QALYs.

These comparisons highlight that while the overall treatment benefit is consistent in direction, the choice of utility weighting principle can influence the estimated magnitude. Sequential captured, QALY related principles generally yield more modest and stable estimates compared to the conventional priciple in this specific case.

4.3 Sensitivity Analysis with Shiny App

Quality-adjusted life year (QALY) utilities can vary depending on the clinical severity of events. To address uncertainty in health state valuations, we adopt default QALY utility values for each event type together with plausible sensitivity ranges, as suggested by Wouters et al [10]. These are summarized in Table 4.3.

For most events, such as myocardial infarction and angina, the sensitivity ranges are relatively narrow. In contrast, stroke demonstrates a wide range of potential utility values, reflecting the heterogeneity of its clinical subtypes and the resulting variability in patients' functional outcomes.

Table 4.3: QALY Utilities and Sensitivity Ranges

Event	**Default**	**Sensitivity Range**
No Event	1.00	–
Cardiovascular Death	0.00	–
Stroke	0.15	0.15 – 0.80
Myocardial Infarction	0.88	0.61 – 0.90
Hospitalizd Angina	0.69	0.69 – 0.90
Revascularization	0.90	0.87 – 1.00

For instance, ischemic strokes with severe sequelae may have a substantial and long-lasting impact on patient well-being, while minor or transient strokes may lead to full recovery. The subtypes of stroke and their corresponding QALY utility values are as follows:

Ischaemic stroke: 0.15

Stroke with severe sequelae: 0.31

Stroke with moderate sequelae: 0.71

Stroke without sequelae: 0.80

An average value of four subtypes above: 0.49

Figure 4.2 displays the QALY-weighted restricted mean survival time (Q-RMST) outcomes under varying stroke utility values with respect to Markov Principle, calculated according to the Markov accumulation principle. The horizontal bars display the Q-RMST for both the intervention group (evolocumab) and the control group (placebo), across each utility assumption. Here we cropped the plot with time interval $[34, 36]$ in order to better visualize the difference between groups.

Although the absolute Q-RMST values shift slightly with changes in the QALY utility assigned to the stroke state, the treatment effect remains consistent across all tested scenarios. Compared to the reference value for ischaemic stroke, the Evolocumab group consistently demonstrates a modest but stable improvement over the control group. The observed difference in Q-RMST ranges from approximately 1.86 to 2.6 days, suggesting that the overall treatment benefit is not highly sensitive to the specific QALY utility attributed to stroke.

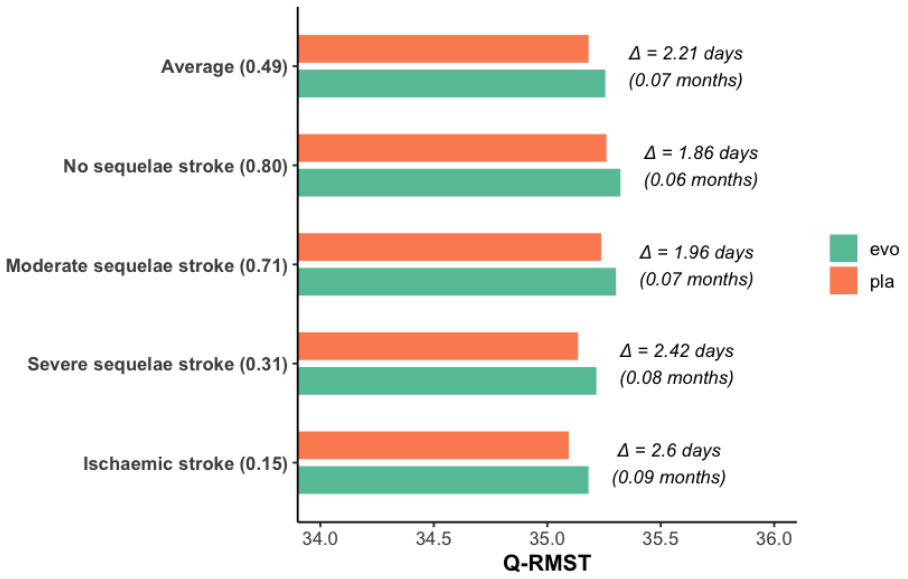

Figure 4.2: Q-RMST Results of different type of stroke under Markov Principle

To facilitate a more comprehensive exploration of this sensitivity and to extend the analysis beyond the stroke state, an interactive Shiny application was developed. This tool allows users to adjust QALY utility values for all clinical events and immediately observe the impact on treatment comparisons. Moreover, it supports toggling between different QALY accumulation principles and enabling users to examine how these modeling assumptions influence Q-RMST outcomes.

Table 4.4: Extreme Q-RMST differences between two groups (in days)

Principle	**Min Diff.**	**Max Diff.**
Markov	1.24	3.37
Worst-state Update	1.75	4.60
Cumulative Product	1.95	5.51
Conventional RMST	9.56	9.56

As presented in Table 4.4, the observed differences in quality-adjusted restricted mean survival time (Q-RMST) between the Evolocumab and Placebo arms vary across QALY accumulation principles and utility weighting scenarios. However, these differences remain relatively modest in magnitude. Under the Markov principle, the Q-RMST difference ranges from 1.24 to 3.37 days, while the Worst-state Update approach yields a slightly broader range of 1.75 to 4.60 days. The Cumulative Product principle shows the greatest sensitivity, with a range from 1.95 to 5.51 days. These variations suggest that while modeling assumptions and utility weights do influence the estimated treatment effect, the overall conclusion - a modest benefit of Evolocumab, remains qualitatively consistent.

In contrast, the conventional RMST approach, which does not incorporate health-state QALY utilities and assumes equal quality of life regardless of clinical events, results in a fixed difference of 9.56 days, regardless of any utility input. This invariance highlights a key limitation of the conventional RMST: it does not reflect differences in patients' quality of life associated with different health states. While it can serve as a benchmark for unadjusted survival time, it lacks the granularity needed to capture the true impact of disease progression and clinical events on patient well-being.

To demonstrate the interactive interface of the Shiny application, we present two representative scenarios corresponding to the minimum and maximum observed Q-RMST differences, based on different combinations of QALY utility values. The application interface consists of two main panels: the left panel features four adjustable sliders, each corresponding to a clinical state (Revascularization, Myocardial Infarction, Hospitalization for Angina, and Stroke), allowing users to assign custom utility values within predefined plausible ranges. The right panel dynamically displays a cropped bar plot comparing Q-RMST differences between the treatment arms across the four modeling principles.

Figure 4.3 illustrates the minimum Q-RMST difference scenario, which occurs when all QALY utility values are set to their upper bounds. In this setting, the perceived impact of events on quality of life is minimized, thereby reducing the difference in quality-adjusted survival time between the two groups. Among all principles, the Markov model yields the smallest treatment difference, with a Q-RMST of only 1.83 days in favor of Evolocumab. This scenario corresponds to the following utility settings: Revascularization = 1.00, Myocardial Infarction = 0.90, Hospitalization for Angina = 0.90, and Stroke = 0.80.

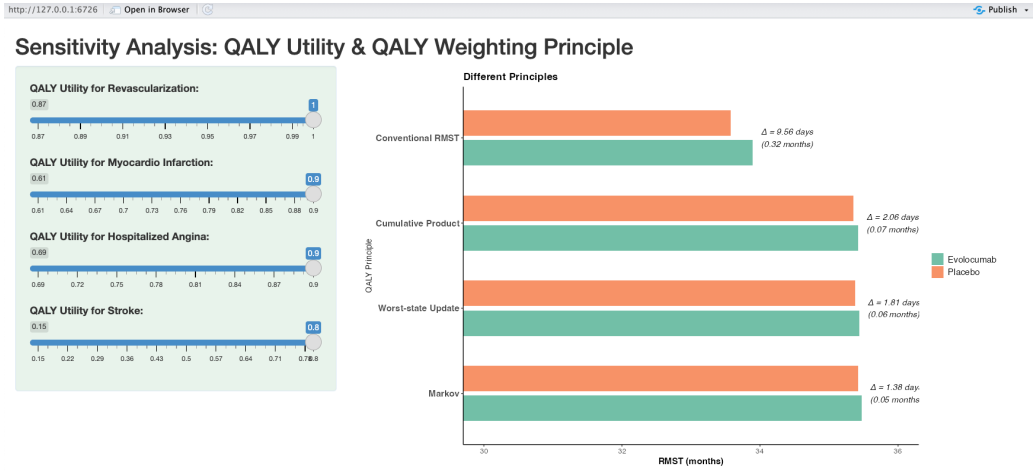

Figure 4.3: Shiny App demonstration of the Minimum Q-RMST difference senario

In contrast, Figure 4.4 presents the maximum Q-RMST difference scenario, in which the utilities of all health states are set to their lower bounds, reflecting a more pessimistic valuation of post-event quality of life. In this configuration, the difference between Evolocumab and Placebo becomes more pronounced across all QALY-based principles. The Cumulative Product method yields the largest observed benefit, with a Q-RMST difference of 5.51 days. This outcome is associated with the following utility values: Revascularization = 0.87, Myocardial Infarction = 0.61, Hospitalization for Angina = 0.69, and Stroke = 0.15.

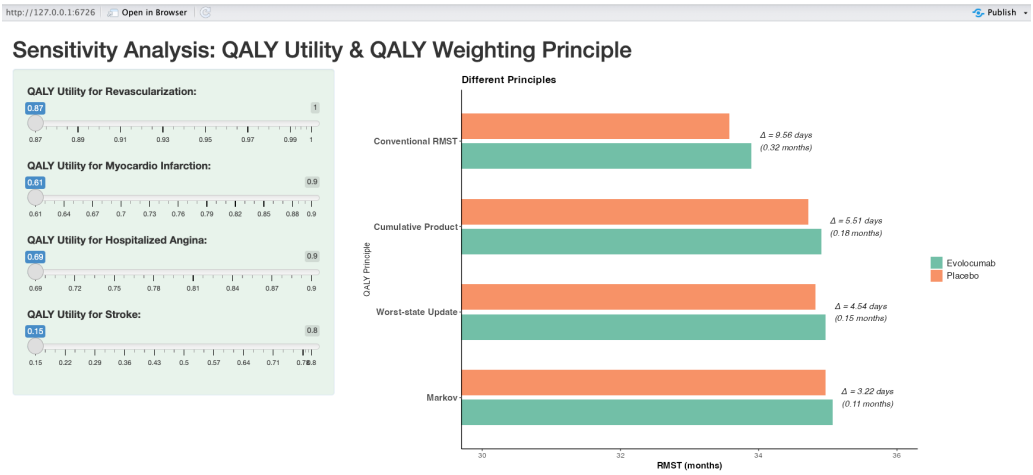

Figure 4.4: Shiny App demonstration of the Maximum Q-RMST difference senario

By enabling users to explore the full range of clinically plausible utility values and observe the resulting changes in treatment effect across different modeling frameworks, the tool supports a more nuanced interpretation of quality-adjusted survival outcomes. With this comprehensive sensitivity analysis complete, we now turn to the conclusion section, where the main findings are summarized.

4.4 Conclusion

Based on the results of the Q-RMST method and supported by the sensitivity analysis, we now conclude the efficacy evaluation of the simulated FOURIER clinical trial data. The Q-RMST analysis yielded a modest and statistically non-significant difference in quality-adjusted survival time between the Evolocumab and Placebo groups. This finding stands in contrast to the original trial's Cox model results, which reported a statistically significant 15% relative reduction in the risk of major cardiovascular events.

The divergence in findings highlights the fundamental differences between traditional time-to-event methods and QALY-based survival metrics. While the Cox model focuses on the timing of first events, the Q-RMST approach captures the full trajectory of a patient's health by accounting for sequential events and incorporating the relative impact of each clinical state on quality of life. Within this framework, although revascularization events are frequent, their limited impact on patients' quality of life results in a relatively small contribution to the overall treatment effect estimate. Additionally, effect sizes of the trial were already low because this is an add-add-on therapy and baseline LDL cholesterol was already close to the zero risk concentration. This shift in weighting may partially explain why the Q-RMST method does not reproduce the statistically significant benefit seen in the original analysis.

It is important to emphasize that these conclusions are based on simulated data and should not be interpreted as a definitive reevaluation of Evolocumab's clinical efficacy. The results presented here are intended to illustrate how alternative modeling frameworks, particularly those that integrate quality of life considerations, can yield different insights. Moreover, this analysis reflects outcomes specific to the assumptions, utility values, and patient population used in the simulation, and may not be generalizable to broader real-world settings.

Nonetheless, this work demonstrates the value of Q-RMST as a complementary approach to traditional survival analysis, offering a patient-centered perspective that aligns better with health economic evaluations and decision-making in clinical practice.

Chapter 5

Discussion

This chapter reflects on the broader implications of the methodology and findings presented in this study. It discusses the applicability and adaptability of the data simulation procedure, the advantages and limitations of the proposed Q-RMST (QALY-weighted Restricted Mean Survival Time) approach, as well as the directions for future work, including potential applications in other fields.

Trial Data

The simulated data used in this study aimed to recreate the multi-state structure and event dynamics of the FOURIER cardiovascular outcomes trial. This allowed for a controlled re-analysis using the Q-RMST framework. However, several assumptions inherent in the simulation process merit further discussion.

We ignored competing deaths from non-cardiovascular causes due to the lack of detailed information on their distribution. However, we note that the overall proportion of non-cardiovascular deaths was relatively low in the context of the full study population. Interestingly, the absolute number of non-cardiovascular deaths was similar to that of cardiovascular deaths. The exact numbers from the paper [6] are presented in Table 5.1:

Table 5.1: Causes of Death in the FOURIER Trial

Outcome	**Evolocumab**	**Placebo**
Death from any cause	444 (3.2%)	426 (3.1%)
Cardiovascular death	*251 (1.8%)*	*240 (1.7%)*
Non-cardiovascular death	*193 (1.4%)*	*186 (1.3%)*

These numbers suggest that while the proportion of non-cardiovascular deaths is modest, their presence remains non-negligible in absolute terms compared to the number of cardiovascular deaths.

A key assumption was the use of constant hazard rates for each event type. While convenient and commonly used in simulations, this assumption may not accurately reflect real-world event dynamics, where hazards often vary over time. To address this, more flexible survival models, such as those based on Weibull, spline-based hazard functions, or piecewise linear functions, could be adopted. These would allow researchers to model time-varying hazards and better capture long-term treatment effects, especially beyond the original follow-up period of the trial.

Another potential improvement for future work lies in the structure of the loss function used to optimize the data simulation. Currently, equal weights are assigned to the components measuring the distance to the summary statistics of the first event, the total number of events, and the distribution of patients by number of events. This uniform weighting leads to a consequence where the difference in the number of first events between the two groups is smaller than reported in the original study, the exact number was showed in Table 2.8 and Table 2.9. That ultimately resulting in a more pessimistic outcome for the later on Q-RMST estimation.

Additionally, the simulation approach has inherent limitations. Most notably, it does not allow for verification of the simulated dataset against actual individual patient data (IPD), since those are not publicly available. Therefore, any inferences drawn from the analysis must be interpreted with caution, as they may not perfectly replicate the real-world distribution of events or patient characteristics.

Q-RMST Method

The Q-RMST framework proposed in this thesis introduces a novel approach to analyzing clinical trial data by integrating survival analysis with quality-adjusted life years (QALYs). A key strength of this method lies in its ability to account for multiple and sequential clinical events, enabling a more comprehensive evaluation of patient trajectories over time. This feature offers a compelling solution to a common issue in clinical trials—the use of composite endpoints, which often combine events of varying clinical importance into a single measure. Such endpoints can obscure the true treatment effect by overemphasizing frequent but less impactful events. In contrast, Q-RMST differentiates between event types based on their timing, order, and associated utility, thereby offering a more patient-centered evaluation of treatment benefit. This represents a significant improvement over traditional methods that consider only time to first event or aggregate event counts, as it better reflects the complexity of disease progression and the real-world experience of patients.

Another advantage is the flexibility to customize QALY utility weights, allowing clinicians and decision-makers to assess the relative burden of different clinical events. However, it is also important to acknowledge that QALY values can vary over time, even within a single health state. Future versions of the model could incorporate time-varying or patient-specific utility values, which would offer a more nuanced view of disease progression and treatment impact.

From an analytical standpoint, the Q-RMST approach can potentially be integrated with existing multi-state modeling frameworks, such as those provided by the `mstate` package in R. This would enable a more principled treatment of censoring and transition probabilities and make the method more accessible to applied researchers.

Clinically, the Q-RMST offers an intuitive interpretation, particularly for patients and non-technical stakeholders. Expressing treatment effects in terms of quality-adjusted survival time is more relatable than relative risk or hazard ratios, which can be abstract and difficult to interpret in meaningful terms. In contrast to measures like odds ratios or hazard ratios, Q-RMST presents absolute differences in time, which often align better with patient priorities and decision-making.

Compared to alternative methods such as the win ratio or win odds—which prioritize event severity and time to event through pairwise comparisons—Q-RMST offers a more comprehensive and continuous perspective. It captures the cumulative impact of multiple events over time and avoids the need for dichotomizing outcomes. As a result, Q-RMST produces an absolute, QALY-utility-based survival metric that is not only clinically informative but also valuable in health economic evaluations.

Application and Future Work

The integration of data simulation with the Q-RMST method offers a promising framework for addressing key limitations of unbalanced composite endpoints in clinical trials. By capturing the timing, sequence, and quality-adjusted impact of multiple events, this approach enhances the interpretability and patient relevance of trial outcomes. Its flexibility allows application across diverse clinical domains, including cardiology, oncology, neurology, and epidemiology, where complex disease trajectories are common. Moreover, the Q-RMST framework aligns naturally with health economic evaluation and cost-effectiveness analyses, where both clinical efficacy and patient-centered outcomes must be considered jointly.

Looking forward, further development could incorporate time-varying or patient-specific QALY utility weights, and deeper integration with multi-state models could improve statistical efficiency and handling of censoring. The method also has potential to inform regulatory and clinical decision-making by providing interpretable, utility-based summaries of treatment benefit.

Moreover, the Shiny app developed for the sensitivity analysis represents an initial step toward creating a tool that bridges the methodology with end users—primarily clinicians. Through its interactive interface, users can adjust the QALY utility values according to their clinical judgment and the specific circumstances of individual patients. They can also explore different principles that might better match how the disease is expected to progress. In addition, the QALY accumulation principles implemented in the tool could be expanded to accommodate a wider range of scenarios in future developments.

Overall, this thesis contributes a meaningful step toward more nuanced, patient-focused evaluation strategies in clinical research, laying the groundwork for future methodological advancements and practical applications in real-world settings.

References

[1] I. Ferreira-González, G. Permanyer-Miralda, J. W. Busse, D. M. Bryant, V. M. Montori, P. Alonso-Coello, S. D. Walter, and G. H. Guyatt, "Methodologic discussions for using and interpreting composite endpoints are limited, but still identify major concerns," *Journal of clinical epidemiology*, vol. 60, no. 7, pp. 651–657, 2007.

[2] S. J. Pocock, C. A. Ariti, T. J. Collier, and D. Wang, "The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities," *European Heart Journal*, vol. 33, no. 2, pp. 176–182, 2012.

[3] J. Butler, N. Stockbridge, and M. Packer, "Win ratio: A seductive but potentially misleading method for evaluating evidence from clinical trials," *Circulation*, vol. 149, no. 20, pp. 1546–1548, May 2024.

[4] C. Ramsay, "Valuing lives and life years: anomalies, implications, and an alternative," *Health Economics, Policy and Law*, vol. 12, no. 3, pp. 337–353, 2017.

[5] Z. R. McCaw, G. Yin, and L.-J. Wei, "Using the restricted mean survival time difference as an alternative to the hazard ratio for analyzing clinical cardiovascular studies," *Circulation*, vol. 140, no. 17, pp. 1366–1368, 2019.

[6] M. S. Sabatine, R. P. Giugliano, A. C. Keech, N. Honarpour, S. D. Wiviott, S. A. Murphy, J. F. Kuder, H. Wang, T. Liu, S. M. Wasserman *et al.*, "Evolocumab and clinical outcomes in patients with cardiovascular disease," *New England journal of medicine*, vol. 376, no. 18, pp. 1713–1722, 2017.

[7] J. Erviti, J. Wright, K. Bassett, M. Ben-Eltriki, C. Jauca, L. C. Saiz, L. Leache, M. Gutiérrez-Valencia, and T. L. Perry, "Restoring mortality data in the fourier cardiovascular outcomes trial of evolocumab in patients with cardiovascular disease: a reanalysis based on regulatory data," *BMJ open*, vol. 12, no. 12, p. e060172, 2022.

[8] N. Liu, Y. Zhou, and J. J. Lee, "IPDfromKM: reconstruct individual patient data from published Kaplan-Meier survival curves," *BMC medical research methodology*, vol. 21, no. 1, p. 111, 2021.

[9] L. C. de Wreede, M. Fiocco, and H. Putter, "mstate: An R Package for the Analysis of Competing Risks and Multi-State Models," *Journal of Statistical Software*, vol. 38, no. 7, pp. 1–30, 2011.

[10] O. J. Wouters, H. Naci, and N. J. Samani, "QALYs in cost-effectiveness analysis: an overview for cardiologists," *Heart*, vol. 101, no. 23, pp. 1868–1873, 2015.

Appendix A

Supplementary Materials

Code Lists

The GitHub repository includes the R code used for data simulation, the Q-RMST method, and the Shiny application. It also contains the corresponding tables and datasets required to run the code. Github repository: Q-RMST

Simulated Data

This is the result of Chapter 2 Data Simulation, contained the data with both evolocumab and placebo group. Simulated Sequential Data

Extracted Puzzle Information Forms

This is some useful information extracted from the Clinical Study Report, which contains the pages with screenshots, and related summary tables of it. Puzzle Information

Estimated constant hazard plots for all types

Estimated constant hazard and cumulative incidence plot of reconstructed type-specific IPD of event stroke and revascularization already provided in the Chapter 3, Section 3.3, here we list the rest types of events: Myocardial Infarction, Hospitalized Angina, and cardiovascular death:

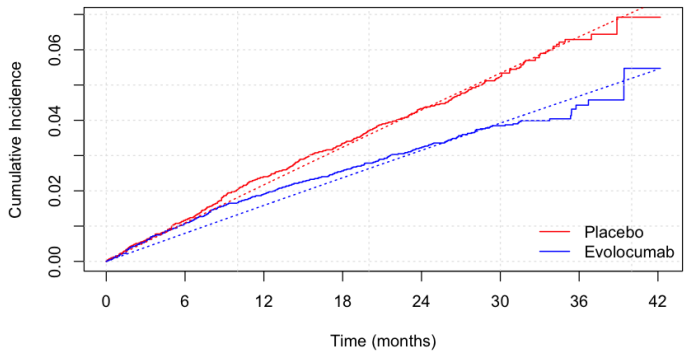

Figure A.1: Estimated constant hazard of event Myocardial Infarction

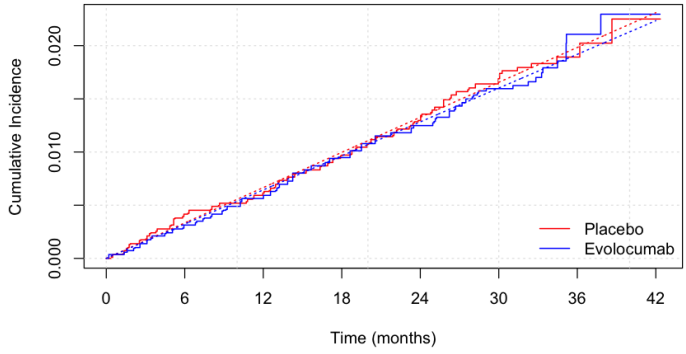

Figure A.2: Estimated constant hazard of event Hospitalized Angina

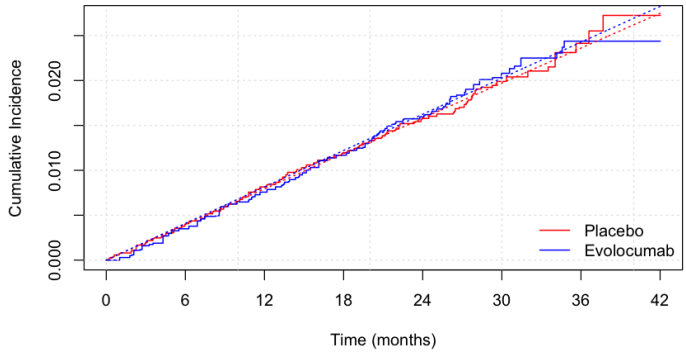

Figure A.3: Estimated constant hazard of event Cardiovascular Death

Appendix B

Use of Generative Artificial Intelligence Tools

During the development of this thesis, I made limited use of generative artificial intelligence tools, specifically OpenAI's ChatGPT, to support the writing and coding process. The assistance provided by the tool was as follows:

1. Writing Support: ChatGPT was used to help rephrase and improve the clarity, coherence, and grammar of selected sections of the thesis. The content and arguments remain my own, and all technical concepts, interpretations, and conclusions were developed independently.

2. Coding Support: The tool was used to review and refine portions of R code used for data simulation, shiny app. In particular, it helped suggest more efficient or readable implementations while preserving the intended functionality.

The use of AI tools did not replace critical thinking, analytical work, or original contributions. All results, methods, and insights reflect my own understanding and work. The tool served as a supplementary aid similar to a writing assistant or code reviewer.