BAYESIAN SURVIVAL ANALYSIS USING DATA FROM ELECTRONIC HEALTH RECORDS: A STUDY ON CARDIOVASCULAR OUTCOMES LEVERAGING INFORMATION FROM RANDOMIZED CLINICAL TRIALS

By

Cara Tanaka Lwin

Master's Thesis

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

in

Biostatistics

May 10, 2024

Nashville, Tennessee

Approved:

Amber J. Hackstadt, Ph.D.

Lauren R. Samuels, Ph.D.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

As the Bayesian approach to data analysis becomes more popular, a common question is what can be gained over traditional methods. One important advantage is that Bayesian analysis allows users to take advantage of previous knowledge about the range and distribution of the parameter values in the model and, therefore, may be advantageous when compared to methods that only take into account the current data set (van de Schoot et al., 2021). Other advantages are a more intuitive interpretation of results and applicability to smaller samples (McNeish, 2016). However, Bayesian methods may be more computationally intensive and are not as widely used, so computational tools can be harder to find or apply. Comparisons of Bayesian and frequentist methods on real data sets can offer additional insights into the benefits and drawbacks of each. This analysis focuses on applying Bayesian survival modeling to a large observational data set obtained from electronic health records (EHR). Utilizing EHR data also brings rewards and challenges (Casey et al., 2016). The large number of covariates present and large sample sizes offer a lot of information but can create computational challenges, and the presence of protected health information limits the type of resources that can be used safely.

Generally, Bayesian statistics assumes that model parameters have an unknown distribution rather than one true value and uses Bayes' Rule to incorporate prior knowledge about this distribution (van de Schoot et al., 2021). The posterior probability is often of interest and is an updated probability that allows the combination of prior knowledge and the likelihood. The influence that the prior has on the posterior depends on how informative it is. In practice, this prior knowledge may come from clinical expertise or previous trials. If there have been many studies and we are confident about the value for that a parameter of interest may take, a highly informative prior with a small variance can be used.

It will not be heavily influenced by new data that is added from the likelihood, and the posterior distribution will strongly resemble the prior distribution. On the other hand, we may be relatively unsure about the values a parameter is most likely to take on and use a less informative or even a non-informative prior. This prior distribution would have a larger variance, and the posterior may be influenced to a greater degree by the new information. The posterior distribution will tend to resemble the likelihood rather than the prior distribution.

In addition to incorporating prior information, Bayesian statistics allows for a more intuitive summary of the result than frequentist methods (Hespanhol et al., 2019). Frequentist confidence intervals are often misunderstood, perhaps because their interpretation depends on hypothetical repeats of the experiment. For example, consider a study that finds that cigarette smokers are 20% more likely to develop pancreatic cancer than nonsmokers, with a 95% confidence interval of 10% to 30%. This means that if the study were conducted 100 times, on average, the confidence interval in 95 of the studies would include the true value. However, because this is never done in real life, it may not be as meaningful or as easy to interpret. Instead of a confidence interval, Bayesian credible intervals offer a very intuitive interpretation: there is a 95% chance that the true value lies within the interval.

The Bayesian posterior predictive distribution is another benefit, as it defines the distribution of future or unobserved values and allows for the prediction of measures that are easily interpreted. For example, in survival analysis, one might be interested in a predicted interval of restricted mean survival time. In the same way that a credible interval is a subset of the parameter space, such that there is a probability α that the parameter is inside the interval, a prediction interval is a subset of the sampling space of the posterior predictive distribution such that there is a probability γ that the value of a future observation is inside the interval. A posterior predictive distribution uses the posterior and considers uncertainty about parameter values, thereby giving an interval that is wider than a confidence interval. However, it is often a more accurate estimate of the range (Lynch and Western, 2004).

This analysis aims to apply Bayesian survival modeling to a large observational data set obtained from electronic health records. The Bayesian approach allows us to incorporate information from previous studies and obtain credible intervals, allowing us to make probability statements when discussing the parameters of interest. To address the lack of randomization, we will implement propensity score matching using the nearest-neighbor approach and a caliper, as it is simple to implement before beginning the Bayesian workflow. We will compare the results of the traditional Cox proportional hazards model to a Bayesian approach using three different priors: one with an uninformative prior, one with a prior derived from a meta-analysis of previous trials, and one with a prior having a small variance. We will compare results by looking at common estimates of interest, including the survival function, hazard ratio, and restricted mean survival time. By comparing these models, we aim to show the effects of various priors and explore the potential benefits of a Bayesian survival analysis approach.

CHAPTER 2

Methods

2.1 Overview

This project extends the analysis done by Richardson et al. in 2023 by considering a Bayesian appreach and incorporating information from other studies. In this new analysis, we perform both the traditional Cox proportional hazards regression analysis and a time-to-event analysis using a Bayesian approach. Sodium-glucose cotransporter-2 inhibitors (SGLT2) and dipeptidyl peptidase 4 inhibitors (DPP4) are classes of diabetes drugs commonly prescribed as add-on therapies to other treatments including metformin, sulfonylurea, and insulin. Their association with major adverse cardiovascular events (MACE) is not well described, especially in populations without pre-existing cardiovascular disease. In this retrospective cohort study of US veterans from 2001-2019, patients were followed from a new prescription fill of a drug in the SGLT2 or DPP4 class until a MACE event, treatment change, loss to follow-up, non-cardiovascular death, or study end (December 2019) (Richardson et al., 2023). A MACE composite outcome was used, consisting of acute myocardial infarction, stroke, heart failure hospitalization, or cardiovascular death. The cohort includes veterans aged 18 years or older with diabetes who were using metformin, sulfonylurea, or insulin alone or in combination. The study focuses on a cohort of diabetes patients without a history of cardiovascular disease. Data was obtained from EHRs and includes inpatient and outpatient VHA encounters, medication fills, and data linkage to Medicare, Medicaid and National Death Index databases. An episode of use was defined with the index date being a prescription of a DPP4 or SGLT2 drug without use of a medication in that class in the prior 180 days, or any other new medication class in the prior 90 days. This wash out period allowed for evaluation of the drug without contamination from a new medication or withdrawal of a different medication (Richardson et al., 2023).

Using this inclusion criteria, the study cohort may include multiple episodes for a single patient.

We consider time-to-event models to model the risk of a MACE event comparing DPP4 and SGLT2. The analysis compares the results from the Cox proportional hazards model, which is a very common frequentist approach to survival analysis, with a Bayesian model estimated using three different priors: one with an uninformative prior that allows the likelihood to have more influence, one with an informative prior that gives strong weight to data from previous studies, and one with an informative prior with small variance to illustrate the strong influence a prior can have on the posterior distribution.

2.2 EHR Studies

The use of data from EHRs has grown quickly over the past decade and provides a low-cost way to utilize valuable longitudinal data on large populations (Casey et al., 2016). In 2009, only 12% of hospitals in the United States reported using EHRs, but this number increased quickly to 76% in 2014 (Charles et al., 2015). This rapid transition has opened up many research possibilities despite EHRs being designed for clinical encounters rather than research needs. Unlike prospective cohort studies which are designed to follow subjects for a given period of time and collect specific variables, EHR studies are limited to what has been collected in the subject's medical records. Even something as simple as cohort definition can be made challenging by the fact that diagnostic codes may be missing or misused. Another challenge is whether conclusions drawn from the EHR population can be generalized to other populations (Hagar et al., 2014). The wealth of data in EHRs makes it worth addressing these challenges. Chart reviews and validation studies can help with improving use of diagnostic codes in the medical record. For example, one study using VHA data to identify fractures found that a modified algorithm increased PPV from 73.5% to 90.1% (Horton et al., 2023). Multiple imputation is often used to address missingness, while propensity score matching can help overcome selection bias inherent in retrospective

cohort studies using EHRs. Weighting to known populations can also be used to address generalizability (Pfeffermann, 1993).

2.3 Propensity Score Matching and Weighting

Because there cannot be randomization in retrospective EHR studies, a crucial part of any observational study is minimizing the selection bias that may be present. Propensity score matching and weighting are two ways to account for differences in observed baseline covariates that may exist between the treatment groups (Benedetto et al., 2018). The propensity score is the conditional probability of assignment to a treatment given a set of covariates (Rosenbaum and Rubin, 1983). Because treatments are not randomly assigned as they would be in a clinical trial, specific characteristics, for example, age, blood pressure, or heart disease, often influence a patient's treatment. While we might not have access to data on all of the factors that determined the treatment, we would like to make use of all of the information available to us. The propensity score is often calculated using a logistic regression model.

$$e(X) = P(Z = 1|X)$$

where $e(X)$ is the propensity score, Z is the exposure with 1 being exposed and 0 being unexposed, and X is a set of observed baseline characteristics (Johnson et al., 2018). Once propensity scores are calculated, there are different ways to use them to balance across treatment groups to minimize confounding, including matching, stratification, inverse probability of treatment weighting (IPTW), and covariate adjustment using the propensity score (Austin, 2011) (Austin and Stuart, 2015).

Propensity score matching is when patients assigned the treatment of interest are matched with a patient in the control group with a similar propensity score. Patients may be matched 1:1 or m:1 between control and treatment groups and also may be matched with or without replacement. In the most commonly used method, nearest neighbor matching, treated

subjects are ordered randomly and are matched one at a time to the control subject with the closest propensity score (Ho et al., 2011). Once a pair is matched, they are removed from the matching pool. Matching continues until all treated subjects are matched or no acceptable matches can be made. Determining whether a match is acceptable is optional and can be specified by setting a maximum difference for the two propensity scores, known as a caliper.

Propensity score weighting, such as inverse probability of treatment weighting (IPTW), has the advantage of including all patients, which usually cannot be achieved with matching. Rather than matching a patient in the treatment group to a patient in the control group, each patient is weighted to create a pseudopopulation in which observed confounders are equally distributed across exposure groups. For example, when using IPTW, each patient in the treatment group is weighted so that patients with a higher probability of receiving treatment have a smaller weight and patients with a lower probability of receiving treatment have a higher weight.

In our analysis, we implemented propensity score matching using the nearest-neighbor approach and a caliper, as it was simple to implement before the Bayesian workflow and adequately balanced the groups. Study covariates included age, sex, race, fiscal year, a surrogate for diabetes duration, diabetes co-therapies, physiologic variables, estimated glomerular filtration rate, summaries of healthcare utilization, smoking, selected comorbidities, and selected medications. After performing mean imputation on missing covariates and adding missingness indicators, 1:1 nearest-neighbor propensity score matching without replacement was used to pair subjects in the treatment (SGLT2) and reference (DPP4) groups. Matching was done using propensity scores estimated using logistic regression with the MatchIt package (Ho et al., 2011). All analysis was done using R and RStudio (R Core Team, 2021) (RStudio Team, 2020)

2.4 Survival Analysis

When the primary outcome is the amount of time until an event of interest occurs, many types of statistical analysis are inappropriate because the event will not occur for some, or perhaps even a majority, of the subjects during the observation period. This leaves us with an unknown time to event. Another feature of this type of data is that it is usually not normally distributed, but rather, events often occur early and then taper off. Several common survival modeling approaches include Kaplan-Meier curves, the log-rank test, Cox proportional hazards regression, and accelerated failure time models (Soodejani et al., 2021).

Kaplan-Meier curves are a non-parametric way to visualize survival data and estimate the survival function. They allow a comparison of the probability of survival at different points in time in the different treatment groups. They are plotted using a non-parametric statistic that estimates the survival function. The probability of surviving longer than time t is given by

$$\hat{S}(t) = \prod_{i:t_i \leq t} (1 - \frac{d_i}{n_i}),$$

where t_i is any time with at least one event, d_i is the number of events at t_i, and n_i is the number of individuals surviving to time t_i. As the sample size increases, this estimator approaches the true survival function (Efron, 1988).

The log-rank test uses a test statistic involving the observed and expected events in each group that can be used to test the null hypothesis that the difference between survival times in the treatment groups is zero. The log-rank statistic as follows has a standard normal distribution:

$$Z = \frac{\sum_{j=1}^{k}(O_j - E_j)}{\sqrt{\sum_{j=1}^{k} V_j}} \sim N(0,1)$$

where O_j is the observed number of events at time t_j, E_j is the expected number of events

if there is no difference between groups, and V_j is the variance of the observed number of events (Wellek, 1993).

Cox proportional hazards regression is beneficial when more data is available, as it directly adjusts for covariates in the hazard function. The hazard is the instantaneous event rate given that a subject has survived to time t. This can help determine the treatment effect while considering factors like age or sex, which is not possible with the basic log-rank test. The Cox model is expressed by the following hazard function,

$$h(t) = h_0(t)\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n)$$

where t is the survival time, $x_1, x_2, ..., x_n$ is a set of n covariates, and $\beta_1, \beta_2, ..., \beta_n$ are the coefficients of the covariates where $\exp(\beta_z)$ is the hazard ratio (HR) of the z^{th} covariate. It can be interpreted such that an HR less than one represents a reduced hazard, and an HR greater than one represents an increased hazard for that covariate. The Cox model is considered semi-parametric because there are no assumptions about the shape of the baseline hazard (Cox, 1972).

Finally, accelerated failure time models are parametric models, commonly based on Weibull or exponential distributions, that explain the effect of covariates as accelerating or decelerating the time-to-event. They also allow for adjustment of covariates, but instead of covariates acting as multipliers of the hazard like in proportional hazards, they act as accelerators or decelerators of the life course of a disease (Wei, 1992).

2.5 Bayesian Survival Modeling

Like the Cox proportional hazards model, the proposed Bayesian approach also uses a hazard model with direct covariate adjustment. We assume the following hazard function for episode p

$$h_p(t) = h_0(t)\exp(\eta_p)$$

where $h_0(t)$ is estimated using M-splines for their ability to adapt to changes in the shape of the hazard function. The linear predictor η_p is expressed as

$$\eta_p = \beta_0 + \beta_1 x_p$$

where x_p is a treatment indicator taking a value of 0 if the episode is in the control group and a value of 1 if the episode is in the treatment group. β_1 is the treatment coefficient where, like the Cox model, $\exp(\beta_1)$ is the hazard ratio of the treatment. Bayesian modeling was performed using the rstanarm R package (Goodrich et al., 2024).

2.5.1 M-splines

M-splines, short for monotone splines, are splines that are non-increasing or non-decreasing functions. In general, a spline is a piecewise polynomial defined on an interval with specified continuity constraints (Ramsay, 1988). The interval is partitioned into subintervals by a knot sequence, k. There is a suitable set of basis splines associated with k, which can be linearly combined to create any other spline associated with this knot sequence (Ramsay, 1988). These monotone piecewise polynomials are useful in capturing a non-linear but monotonic relationship, meaning that the hazard is either non-increasing or non-decreasing with the predictor variable. They are very flexible and allow approximation of complex functions by piecing together simpler polynomial functions in a smooth and continuous manner. In terms of the baseline hazard, the M-splines model can be expressed in the following way:

$$h_p(t) = \sum_{l=1}^{L} \gamma_l M_l(t; \mathbf{k}, \delta) \exp(\eta_p)$$

where $M_l(t; \mathbf{k}, \delta)$ denotes the l^{th} basis term for a degree δ M-spline function evaluated at a vector of knot locations $\mathbf{k} = \{k_1, ..., k_J\}$ and γ_l is the l^{th} M-spline coefficient (Brilleman et al., 2020). Denote the M-spline for the baseline hazard $\sum_{l=1}^{L} \gamma_l M_l(t; \mathbf{k}, \delta)$ as $M(t; \gamma, \mathbf{k}, \delta)$.

The I-spline evaluated using the same degree δ, knot locations $\mathbf{k}$, and coefficients γ is the integral of the M-spline and is denoted as $I(t;\gamma,\mathbf{k},\delta)$ (Wang and Yan, 2018). The parameterizations of the hazard function, cumulative hazard function, survival function, and cumulative incidence function (CIF) are as follows (Brilleman et al., 2020).

$$h_p(T_p) = M(T_p;\gamma,\mathbf{k},\delta)\exp(\eta_p)$$

$$H_p(T_p) = I(T_p;\gamma,\mathbf{k},\delta)\exp(\eta_p)$$

$$S_p(T_p) = \exp(-I(T_p;\gamma,\mathbf{k},\delta)\exp(\eta_p))$$

$$CIF_p(T_p) = 1-\exp(-I(T_p;\gamma,\mathbf{k},\delta)\exp(\eta_p))$$

The hazard function defines the instantaneous rate of the event at time T_p given that an individual has survived up to that time, and the cumulative hazard function is the integral of the hazard function, which represents the total hazard experienced up to time T_p. The survival function represents the probability that an individual survives past time T_p, while the CIF is the complement of the survival function and defines the probability that the event time is less than T_p.

2.5.2 Specification of Priors

When information is known about the location or scale of the hazard ratio, a prior can be placed on the corresponding β of the linear predictor, η_p, defined in section 2.5. While many different distributions can be specified, the normal distribution is a straightforward way to add existing information about the mean and standard deviation where the mean is the log of the known HR.

2.5.3 Posterior Distribution

The posterior is proportional to the likelihood times the prior and often does not have a closed-form expression. Markov chain Monte Carlo (MCMC) methods provide a way to

sample from the posterior without requiring it to have an explicit form. To simplify greatly, a Markov chain is a sequence of parameter values where each value is generated based on the previous one, and the chain eventually converges to the desired posterior distribution (Roberts and Rosenthal, 2004). Hamiltonian Monte Carlo (HMC) is an MCMC algorithm that integrates concepts from Hamiltonian mechanics to guide the exploration of the target distribution (Hoffman and Gelman, 2014). HMC takes into account the underlying geometry of the distribution to make exploration of the parameter space more efficient and also results in faster convergence compared to traditional random-walk MCMC methods because it tends to produce less correlated samples (Almond, 2014).

2.5.4 Estimates of Interest

In time-to-event studies, values of interest include the hazard ratio and the restricted mean survival time. Hazard ratios compare the instantaneous probability that an individual experiences the event in the treatment group versus the control group (Spruance et al., 2004). For example, a hazard ratio less than one means that in a group of subjects who have not experienced the event, a subject in the treatment group is less likely to have an event in the next time period compared to a control. The posterior for the HR associated with coefficient x_i can be found by exponentiating posterior draws from the posterior for β_i. However, the HR alone does not determine treatment benefit, as the shape of the underlying probability distribution also plays a role. There are cases in which a hazard ratio far from one does not result in a large change in survival time. The restricted mean survival time gives us another way to look at treatment benefit, as it represents the average survival time during a defined time period. It is a helpful way to translate the results to a metric that is easy to understand. As an example, to determine the RMST at 5 years, we would integrate the area under the survival curve for each treatment up to the 5-year mark. If treatment A has an RMST of 3.8 years while treatment B has an RMST of 4.6 years, it would suggest that treatment B is associated with better survival outcomes within the 5-year time frame. This

provides a straightforward measure of treatment benefit over a specific time interval. The RMST is found using the posterior predictive distribution for the survival function using an established R function (Elçi and Brilleman, 2019).

CHAPTER 3

Results

3.1 Propensity Score Matching

In the retrospective cohort, 129,834 episodes of DPP4 and SGLT2 use meet the inclusion criteria. Of the 129,834 episodes, 23,107 used an SGLT2 (treatment) drug, while 106,727 used a DPP4 (reference) drug. After matching using a caliper value of 0.05, 21,821 episodes from the treatment group were matched, and 1,286 were dropped. After matching, all standardized mean differences of observed baseline covariates were below the acceptable threshold of 0.1. The patient characteristics of the matched cohort are summarized in Table 3.1 and Table 3.2.

3.2 Survival Analysis

The methods we compare are a Cox proportional hazards model (Cox PH), a Bayesian survival model with an uninformative prior (Uninformative prior), a Bayesian survival model with small variance (Small variance prior), and a Bayesian survival model with a prior derived from a Zelniker et al. meta-analysis that combined data from three cardiovascular outcome trials (Zelniker prior), finding a hazard ratio of 0.86 (0.80-0.93) (Zelniker et al., 2019). This was implemented using a normal (log(0.86), 0.0385) prior on the treatment coefficient. Although the populations represented in the meta-analysis differ from our population in that subjects had cardiovascular disease, it represents many clinicians' prior belief that there is a protective effect of SGLT2 on MACE outcomes. The small variance prior is meant to show what happens to our estimates and their credible intervals when the prior is very strong. We call the uninformative prior as such because of its relatively flat shape, although it is actually a normal distribution with a large variance. Its shape and small amount of influence on the posterior can be compared to the Zelniker prior in Figure 3.1.

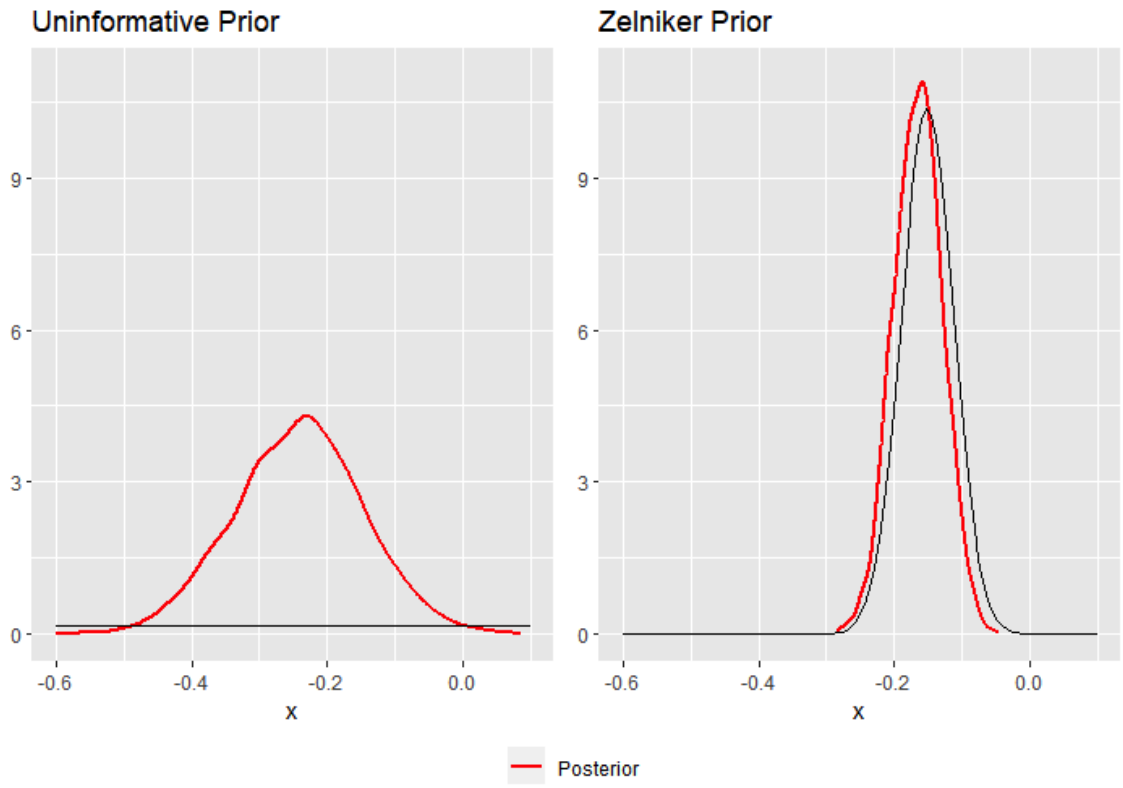

Figure 3.1: The posterior distribution plotted in red with its prior distribution plotted in black for two of the Bayesian survival models

3.2.1 Posterior distributions

The influence of the three different prior distributions can be seen in the location and shape of the posteriors of the β on the treatment variable (Figure 3.2). The posterior resulting from the uninformative prior is relatively wide, and its location was not highly influenced by the prior. In contrast, the posterior distribution using the Zelniker prior was pulled toward the location of the prior and also had a smaller range. Finally, the posterior using the small variance prior is completely centered on the location of the prior with an extremely small range.

3.2.2 Hazard Ratio

We estimate the hazard ratio for SGLT2 with DPP4 as the reference group for each model (Section 2.5.4). The Bayesian model with uninformative prior resulted in a hazard ratio

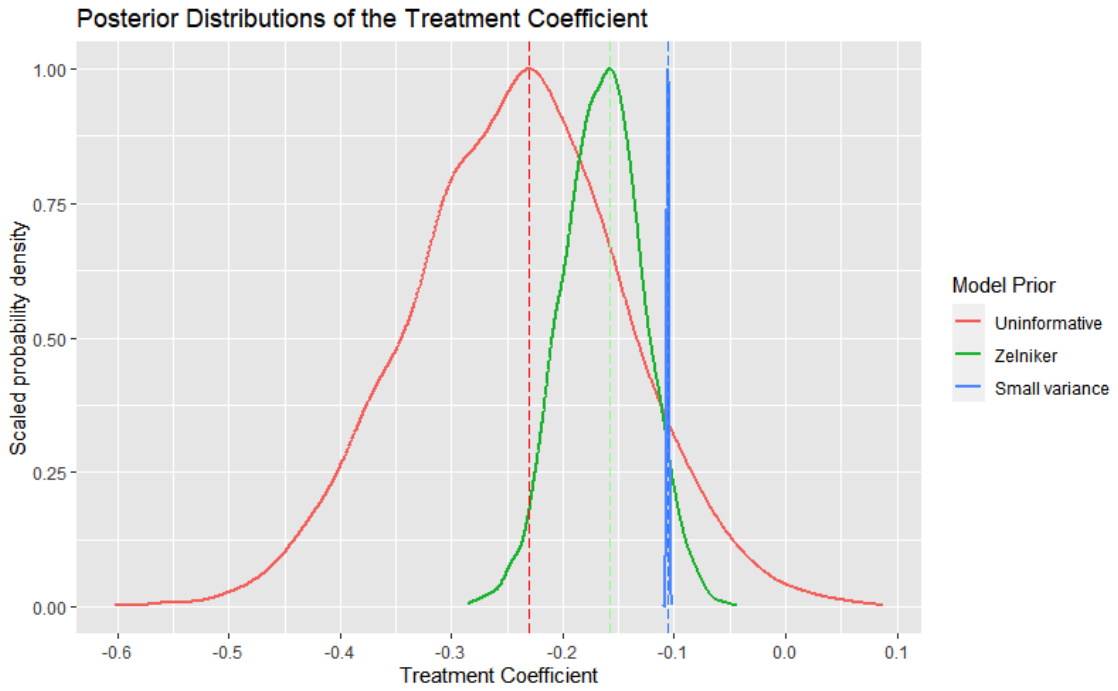

Figure 3.2: Comparison of posterior distributions of the treatment coefficient using 3 different priors with the estimate of the treatment coefficient marked with a dashed line

estimate of 0.787 and 95% credible interval of (0.651 - 0.946), which is very similar to the Cox proportional hazards model estimate of 0.784 with 95% confidence interval (0.652 - 0.943) (Table 3.4). The 95% credible interval is found using the 0.025 and 0.975 percentile of the posterior distribution and is an equal-tailed interval. The 95% credible interval for the Bayesian analysis using an uninformative prior was also very similar to the confidence interval of the Cox PH model. Using the Zelniker prior, the hazard ratio is estimated to be 0.849 with a 95% probability of being between 0.791 and 0.912. The HR increased toward the mean of the prior, and the credible interval was also much narrower compared to the model with the uninformative prior. The small variance prior pulls the HR estimate up to the mean of the prior distribution with an HR estimate of 0.90 and creates an extremely narrow credible interval of 0.898 to 0.902. All models estimate a protective effect of SGLT2 on MACE outcomes.

3.2.3 Survival Function

We estimate the survival function for SGLT2 versus DPP4 using the three Bayesian survival models (uninformative, Zelniker, and small-variance). The survival plot of the uninformative prior with the hazard ratio estimate farthest from one has the largest separation between survival curves in the two groups (Figure 3.3). There is some overlap of the credible intervals, especially after about 2.5 years. The model using the Zelniker prior has more overlap of the pointwise credible intervals. Finally, the model with a small-variance prior estimates the hazard ratio closest to one out of the three models and its survival functions and pointwise credible intervals strongly overlap.

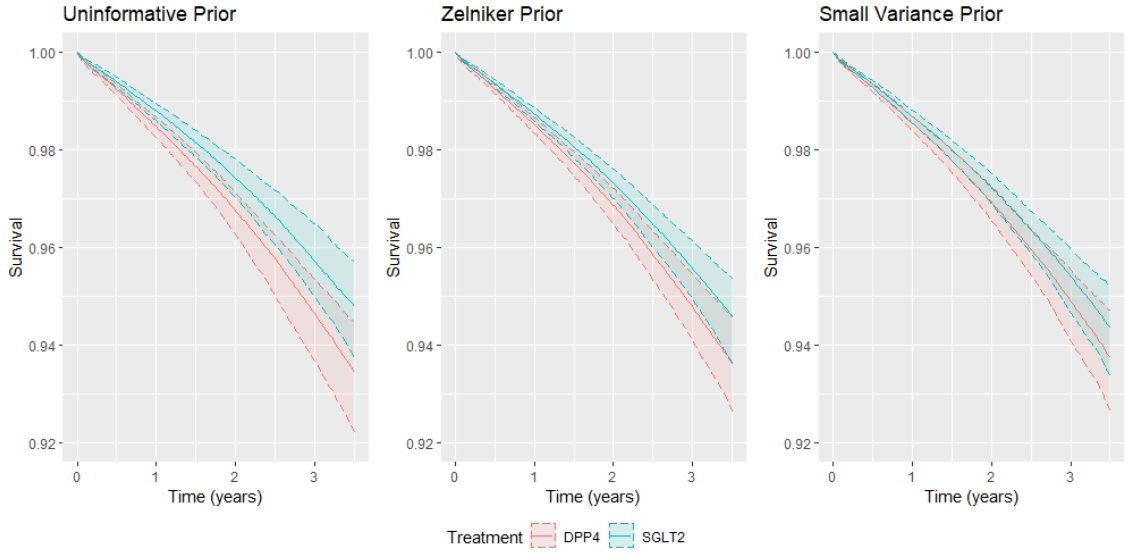

Figure 3.3: Survival functions of three Bayesian survival models

3.2.4 Restricted Mean Survival Time

There is a similar pattern to the survival functions when comparing restricted mean survival time at three years found using the frequentist approach and the Bayesian survival model with specified priors (Figure 3.4). Starting with the uninformative prior with the lowest hazard ratio estimate, the difference in estimated RMST is about 0.016 years between treatment groups (Table 3.5). This is similar to the RMST derived from the Kaplan Meier estimate of the survival function, which had a difference in estimated RMST of 0.015

years. Moving to the model with the Zelniker prior, the hazard ratio estimate increases, and the difference in estimated RMST decreases to 0.011 years. Lastly, the model with the small variance prior has a difference in estimated RMST of 0.007 years between treatment groups. As expected, when the model estimates less difference between treatments, predictions like restricted mean survival time also become closer together. Like the KM function, all three Bayesian models estimate a larger RMST for SGLT2 users, though the credible intervals overlap, so no strong conclusions can be drawn about the difference in RMST between the two treatment groups (Table 3.5).

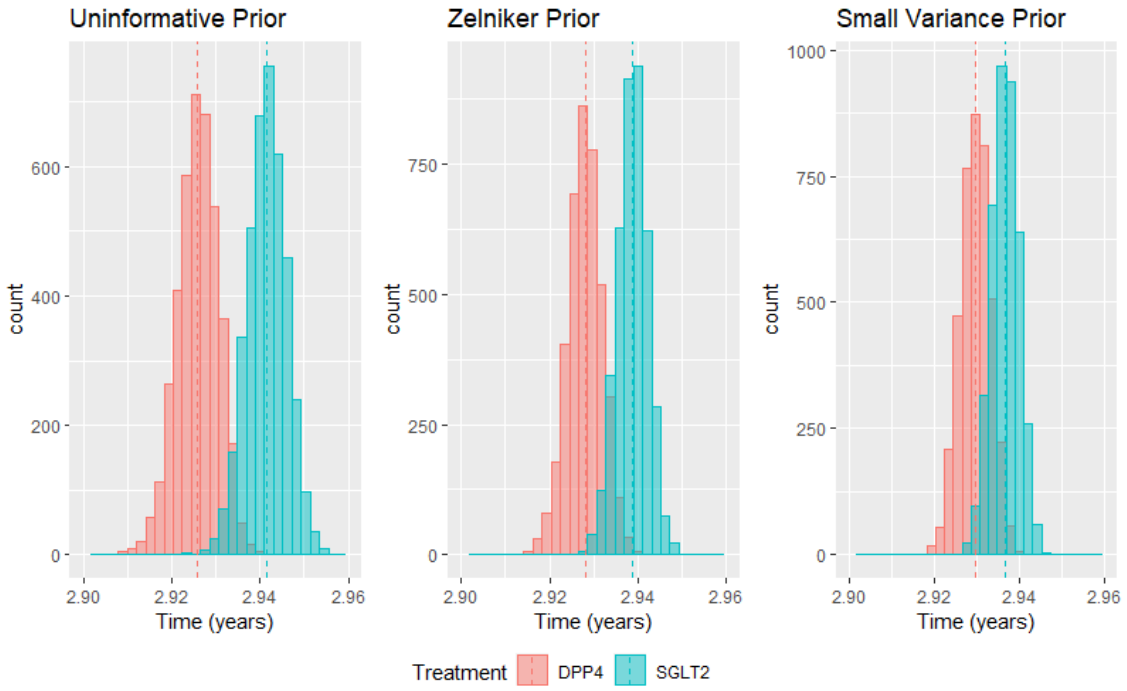

Figure 3.4: Comparison of the distributions of the restricted mean survival time (in years) at 3 years using three Bayesian models

	DPP4	SGLT2	SMD
n	21822	21822	
Age (mean (SD))	65.29 (11.00)	65.12 (9.68)	0.017
Male, N (%)	20536 (94.1)	20572 (94.3)	0.007
Race, N (%)			0.003
Other	808 (3.9)	799 (3.9)	
Black or African American	4553 (22.0)	4567 (22.1)	
White	15310 (74.1)	15342 (74.1)	
Race Missing (%)	1151 (5.3)	1114 (5.1)	0.008
VISN (%)			0.037
1	900 (4.1)	917 (4.2)	
2	673 (3.1)	686 (3.1)	
3	67 (0.3)	64 (0.3)	
4	1080 (4.9)	1059 (4.9)	
5	1019 (4.7)	1075 (4.9)	
6	1275 (5.8)	1287 (5.9)	
7	1701 (7.8)	1701 (7.8)	
8	1864 (8.5)	1916 (8.8)	
9	885 (4.1)	860 (3.9)	
10	1489 (6.8)	1495 (6.9)	
11	88 (0.4)	99 (0.5)	
12	640 (2.9)	654 (3.0)	
15	1146 (5.3)	1075 (4.9)	
16	1656 (7.6)	1601 (7.3)	
17	2136 (9.8)	2156 (9.9)	
18	145 (0.7)	156 (0.7)	
19	1234 (5.7)	1125 (5.2)	
20	1136 (5.2)	1153 (5.3)	
21	732 (3.4)	720 (3.3)	
22	1095 (5.0)	1110 (5.1)	
23	861 (3.9)	913 (4.2)	
Date (mean (SD))	17717.84 (460.30)	17716.38 (486.89)	0.003
Diabetes medication start to index date (mean (SD))	3463.43 (1850.18)	3474.01 (1831.25)	0.006
Height (mean (SD))	69.55 (3.11)	69.60 (3.08)	0.015
Height Missing (%)	4690 (21.5)	4631 (21.2)	0.007
Weight (mean (SD))	232.01 (51.08)	232.67 (48.21)	0.013
Weight Missing (%)	453 (2.1)	431 (2.0)	0.007
BMI (mean (SD))	33.70 (6.85)	33.76 (6.41)	0.010
BMI Missing (%)	4966 (22.8)	4895 (22.4)	0.008
Systolic blood pressure, mm/Hg (mean (SD))	134.29 (16.54)	134.22 (16.33)	0.005
Diastolic blood pressure mm/Hg (mean (SD))	76.43 (9.90)	76.46 (9.78)	0.003
Blood pressure measure missing (%)	417 (1.9)	391 (1.8)	0.009
Hemoglobin, g/dL (mean (SD))	14.24 (2.06)	14.24 (1.62)	0.002
Hemoglobin Missing (%)	1906 (8.7)	1862 (8.5)	0.007
Estimated glomerular filtration rate, ml/min (mean (SD))	80.72 (20.98)	80.79 (18.61)	0.003
eGFR missing (%)	1278 (5.9)	1222 (5.6)	0.011
Low density lipoprotein (mean (SD))	86.88 (34.88)	86.67 (34.94)	0.006
ow density lipoprotein missing (%)	1269 (5.8)	1222 (5.6)	0.009
HbA1c (mean (SD))	8.70 (1.63)	8.71 (1.54)	0.007
HbA1c missing (%)	1648 (7.6)	1574 (7.2)	0.013
Urine protein on urinalysis (%)			0.016
Null unknown or Negative	8965 (41.1)	9132 (41.8)	
Trace or 1+	2113 (9.7)	2100 (9.6)	
2+	1201 (5.5)	1185 (5.4)	
3+/4+/trace to 4+	241 (1.1)	239 (1.1)	
Missing	9302 (42.6)	9166 (42.0)	
Microalbumin to creatine ratio (%)			0.011
A1 and unknown but tested	9416 (43.1)	9421 (43.2)	
A2	4102 (18.8)	4186 (19.2)	
A3 and positive	1414 (6.5)	1405 (6.4)	
Missing	6890 (31.6)	6810 (31.2)	

Table 3.1: Baseline covariates in the matched sample: Demographics and lab values, with standardized mean difference (SMD)

	DPP4	SGLT2	SMD
n	21822	21822	
Malignancy (%)	2220 (10.2)	2187 (10.0)	0.005
Liver disease (%)	1174 (5.4)	1180 (5.4)	0.001
HIV (%)	95 (0.4)	103 (0.5)	0.005
Congestive heart failure (%)	1034 (4.7)	1077 (4.9)	0.009
Serious mental illness (%)	6443 (29.5)	6471 (29.7)	0.003
Smoking (%)	1832 (8.4)	1873 (8.6)	0.007
Chronic obstructive pulmonary disease (%)	2811 (12.9)	2773 (12.7)	0.005
History of respiratory failure (%)	490 (2.2)	499 (2.3)	0.003
History of kidney disease (%)	0 (0.0)	1 (0.0)	0.010
History of sepsis (%)	268 (1.2)	271 (1.2)	0.001
History of pneumonia (%)	303 (1.4)	325 (1.5)	0.008
Arrhythmias (%)	992 (4.5)	982 (4.5)	0.002
Cardiac valve disease (%)	358 (1.6)	340 (1.6)	0.007
Parkinson's (%)	161 (0.7)	149 (0.7)	0.007
Urinary tract infection (%)	648 (3.0)	636 (2.9)	0.003
Osteomyelitis (%)	115 (0.5)	104 (0.5)	0.007
Osteoporosis (%)	109 (0.5)	134 (0.6)	0.015
Falls (%)	179 (0.8)	200 (0.9)	0.010
Fractures (%)	328 (1.5)	322 (1.5)	0.002
Amputation (%)	59 (0.3)	61 (0.3)	0.002
Retinopathy (%)	2287 (10.5)	2366 (10.8)	0.012
ACE inhibitors (%)	11413 (52.3)	11383 (52.2)	0.003
Angiotensin receptor blockers (%)	5130 (23.5)	5111 (23.4)	0.002
Beta blockers (%)	7998 (36.7)	8035 (36.8)	0.004
Calcium channel blockers (%)	6633 (30.4)	6614 (30.3)	0.002
Thiazide/potassium sparing diuretics (%)	7158 (32.8)	7195 (33.0)	0.004
Loop diuretics (%)	2400 (11.0)	2397 (11.0)	<0.001
Other hypertensives (%)	5655 (25.9)	5631 (25.8)	0.003
Lipid-lowering statins (%)	17235 (79.0)	17253 (79.1)	0.002
Non-statin lipid-lowering agents (%)	3067 (14.1)	3105 (14.2)	0.005
Anti-arrhythmic digoxin and inotropes (%)	1257 (5.8)	1222 (5.6)	0.007
Anticoagulants (%)	1814 (8.3)	1772 (8.1)	0.007
Nitrates (%)	679 (3.1)	660 (3.0)	0.005
Aspirin (%)	4381 (20.1)	4420 (20.3)	0.004
Platelet inhibitors (%)	931 (4.3)	907 (4.2)	0.005
Antipsychotics (%)	1317 (6.0)	1317 (6.0)	<0.001
Oral glucocorticoids (%)	2069 (9.5)	1979 (9.1)	0.014
Hospitalization within year (Veterans Health) (%)	1274 (5.8)	1336 (6.1)	0.012
Hospitalization within 30 days (Veterans Health) (%)	208 (1.0)	224 (1.0)	0.007
Hospitalization within year (Medicaid/Medicare) (%)	634 (2.9)	649 (3.0)	0.004
Hospitalization within 30 days (Medicaid/Medicare) (%)	72 (0.3)	75 (0.3)	0.002
Medicaid insurance use in last year	248 (1.1)	255 (1.2)	0.003
Medicare insurance use in last year	8586 (39.3)	8178 (37.5)	0.038
Medicare advantage use	0.01 (0.11)	0.01 (0.12)	0.004
Nursing home encounters	48 (0.2)	50 (0.2)	0.002
Cotherapy (%)			0.022
Insulin	2982 (13.7)	2980 (13.7)	
Metformin	3662 (16.8)	3569 (16.4)	
Metformin + Insulin	6407 (29.4)	6603 (30.3)	
Metformin + Sulfonylurea	6827 (31.3)	6719 (30.8)	
Sulfonylurea	1177 (5.4)	1170 (5.4)	
Sulfonylurea + Insulin	767 (3.5)	781 (3.6)	
Ejection fraction at baseline			0.017
Indeterminate	535 (10.6)	535 (10.5)	
Missing/Unknown	41 (0.8)	42 (0.8)	
Normal/Inc	3908 (77.7)	3916 (77.2)	
Reduced/Severe	548 (10.9)	579 (11.4)	
Ejection fraction at baseline missing	16790 (76.9)	16750 (76.8)	0.004
Outpatient visits in last year (mean (SD))	6.95 (6.33)	7.09 (6.51)	0.022
Number of medications	4.56 (2.40)	4.56 (2.35)	0.001

Table 3.2: Baseline covariates in the matched ssample: Comorbidities and prescriptions, with standardized mean difference (SMD)

Method	Coefficient Estimate	Lower	Upper
Cox PH [1]	-0.243	-0.429	-0.057
Uninformative prior [2]	-0.239	-0.430	-0.055
Zelniker prior [2]	-0.163	-0.235	-0.095
Small variance prior [2]	-0.105	-0.107	-0.103

Table 3.3: A comparison of treatment coefficient estimates using Cox PH and Bayesian survival model with uninformative prior, Zelniker prior, and small variance prior

[1]Interval shown is a 95% confidence interval

[2] Interval shown is a 95% credible interval.

Method	HR Estimate	Lower	Upper
Cox PH [1]	0.784	0.652	0.943
Uninformative prior [2]	0.787	0.651	0.946
Zelniker prior [2]	0.849	0.791	0.912
Small variance prior [2]	0.900	0.898	0.902

Table 3.4: A comparison of hazard ratio estimates using a Cox proportional hazards method, Bayesian survival model with uninformative prior, and Bayesian survival model with Zelniker prior

[1]Interval shown is a 95% confidence interval

[2] Interval shown is a 95% credible interval.

Method	Treatment	RMST	Lower	Upper
Cox PH [1]	DPP4	2.926	2.916	2.935
	SGLT2	2.941	2.931	2.950
	Difference	0.015	0.001	0.028
Uninformative prior [2]	DPP4	2.926	2.916	2.934
	SGLT2	2.942	2.933	2.950
	Difference	0.016	0.004	0.028
Zelniker prior [2]	DPP4	2.928	2.920	2.935
	SGLT2	2.939	2.932	2.945
	Difference	0.011	0.006	0.016
Small variance prior [2]	DPP4	2.930	2.923	2.936
	SGLT2	2.937	2.930	2.943
	Difference	0.007	0.006	0.008

Table 3.5: A comparison of restricted mean survival time (in years) at 3 years.

[1]RMST estimated using Kaplan Meier estimate for the survival function; interval shown is a 95% confidence interval.

[2] Interval shown is a 95% credible interval.

CHAPTER 4

Discussion

The results from all of the Bayesian models considered suggest a potential protective effect of SGLT2 on MACE outcomes. In addition to using an uninformative prior, as is used for many Bayesian models, the Bayesian survival model allowed us to incorporate information from previous studies using informative priors. The Zelniker prior is an example of incorporating information about the hazard ratio of 0.80-0.93 found in other studies, although the population in the Zelniker meta-analysis differs in that the population had existing cardiovascular disease. The relatively small confidence interval on the Zelniker estimate brings the Bayesian HR estimate much closer to the prior and makes the credible interval smaller. Given the differences in populations, one could also consider a prior weighting to a HR of 0.86, but with a larger variance. The small variance prior illustrates just how influential a prior can be, as the posterior distribution essentially follows the prior distribution. Clearly, the choice of variance in the prior is important to consider when determining how much influence to give the prior information, and caution should be used when implementing normal priors with small variance. Overall, the finding of a protective SGLT2 treatment effect is important as many existing studies use populations with existing cardiovascular disease, while fewer studies have been published on those without CVD. The Bayesian approach allowed us to utilize the large amount of information in the EHR data set and include information from other randomized trials.

While there are challenges when using data from retrospective studies and EHRs, there are well-developed methods like propensity score matching, imputation of missing values, and careful data validation to help address them. We incorporated propensity score matching with Bayesian models to help adjust for the lack of randomization in EHR studies, although we are limited to adjusting for observed baseline covariates. Using data from the

Veterans Health Administration adds another potential drawback because it tends to have an over-representation of White and male subjects as compared to the U.S. population. However, VHA data is especially rich, with information on demographics, diagnostics, and procedures, as well as linkage to Medicare and Medicaid prescription data that makes it a useful source of information despite its lack of generalizability to certain patient populations.

It is possible for multiple episodes in the study cohort to be from the same patient (Richardson et al., 2023). The current Bayesian approach does not account for the correlation between multiple episodes from the same patient. However, the estimate for the uninformative hazard ratio is very similar to the result in the Richardson 2023 analysis that incorporates this correlation into the model. An approach using weighting, as opposed to propensity score matching, can include an adjustment for multiple episodes in the weights.

Given the similarities in the HR estimate between the traditional Cox proportional hazards model and the Bayesian survival model with an uninformative prior, the question arises of when to use a Bayesian approach. Even without prior clinical knowledge, a Bayesian model with an uninformative prior is useful when credible intervals or predictive distributions are desired. The addition of clinical information or estimates from previous trials adds another benefit with the ability to incorporate that knowledge using informative priors. Although Bayesian methods can be computationally intensive, several packages now exist that make the analysis relatively straightforward and worthwhile. We found Bayesian survival models with PS matching to be a useful tool for incorporating information from previous trials into our EHR study. Possible next steps include utilizing propensity score weighting, performing data-driven variable selection for propensity score and outcome models, and the development of more efficient algorithms for large EHR data sets.

References

Almond, R. (2014). A comparison of two MCMC algorithms for hierarchical mixture models.

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*.

Austin, P. C. and Stuart, E. A. (2015). Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Statistics in Medicine*.

Benedetto, U., Head, S. J., Angelini, G. D., and Blackstone, E. H. (2018). Statistical primer: propensity score matching and its alternatives. *European Journal of Cardio-Thoracic Surgery*.

Brilleman, S. L., Elci, E. M., Novik, J. B., and Wolfe, R. (2020). Bayesian survival analysis using the rstanarm R package.

Casey, J. A., Schwartz, B. S., Stewart, W. F., and Adler, N. E. (2016). Using electronic health records for population health research: A review of methods and applications. *Annual Review of Public Health*.

Charles, D., Gabriel, M., and Searcy, T. (2015). Adoption of electronic health record systems among u.s. nonfederal acute care hospitals: 2008-2014. *ONC Data Brief*.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society*.

Efron, B. (1988). Logistic regression, survival analysis, and the kaplan-meier curve. *Journal ofo the American Statistical Association*.

Elçi, E. M. and Brilleman, S. (2019). Bayesian analyses of time-to-event data using the rstanarm R package. StanCon conference paper.

Goodrich, B., Gabry, J., Ali, I., and Brilleman, S. (2024). rstanarm: Bayesian applied regression modeling via Stan. R package version 2.32.1.

Hagar, Y., Albers, D., Pivovarov, R., Chase, H., Dukic, V., and Elhadad, N. (2014). Survival analysis with electronic health record data: Experiments with chronic kidney disease. *Statistical Analysis and Data Mining*.

Hespanhol, L., Vallio, C. S., Costa, L. M., and Saragiotto, B. T. (2019). Understanding and interpreting confidence and credible intervals around effect estimates. *Brazilian Journal of Physical Therapy*.

Ho, D. E., Imai, K., King, G., and Stuart, E. A. (2011). MatchIt: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*, 42(8):1–28.

Hoffman, M. D. and Gelman, A. (2014). The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*.

Horton, T. G., Jr., T. L. R., Hackstadt, A. J., Halvorson, A. E., Hung, A. M., Greevy, R., and Roumie, C. L. (2023). Validation of an algorithm to identify fractures among patients within the veterans health administration. *Pharmacoepidemiology and Drug Safety*.

Johnson, S. R., Tomlinson, G. A., Hawker, G. A., Granton, J. T., and Feldman, B. M. (2018). Propensity score methods for bias reduction in observational studies of treatment effect. *Rheumatic Disease Clinics of North America*.

Lynch, S. M. and Western, B. (2004). Bayesian posterior predictive checks for complex models. *Sociological Methods and Research*.

McNeish, D. (2016). On using bayesian methods to address small sample problems. *Structural Equation Modeling*.

Pfeffermann, D. (1993). The role of sampling weights when modeling survey data. *International Statistial Review*.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ramsay, J. O. (1988). Monotone regression splines in action. *Statistical Science*.

Richardson, T. L., Halvorson, A. E., Hackstadt, A. J., Hung, A. M., Greevy, R., Grijalva, C. G., Elasy, T. A., and Roumie, C. L. (2023). Primary occurrence of cardiovascular events after adding sodium–glucose cotransporter-2 inhibitors or glucagon-like peptide-1 receptor agonists compared with dipeptidyl peptidase-4 inhibitors: A cohort study in veterans with diabetes. *Annals of Internal Medicine*.

Roberts, G. O. and Rosenthal, J. S. (2004). General state space markov chains and MCMC algorithms. *Probability Surveys*.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*.

RStudio Team (2020). *RStudio: Integrated Development Environment for R*. RStudio, PBC., Boston, MA.

Soodejani, M., Tabatabaei, S., and Mahmoudimanesh, M. (2021). Bayesian statistics versus classical statistics in survival analysis: an applicable example. *American Journal of Cardiovascular Disease*.

Spruance, S. L., Reid, J. E., Grace, M., and Samore, M. (2004). Hazard ratio in clinical trials. *Antimicrobial Agents and Chemotherapy*.

van de Schoot, R., Depaoli, S., King, R., Kramer, B., Martens, K., Tadesse, M. G., Vannucci, M., Gelman, A., Veen, D., Willemsen, J., and Yau, C. (2021). Bayesian statistics and modelling. *Nature Reviews*.

Wang, W. and Yan, J. (2018). splines2: Regression spline functions and classes. R package version 0.2.8.

Wei, L. J. (1992). The accelerated failure time model: A useful alternative to the cox regression model in survival analysis. *Statistics in Medicine*.

Wellek, S. (1993). A log-rank test for equivalence of two survivor functions. *International Biometric Society*.

Zelniker, T. A., Wiviott, S. D., Raz, I., Im, K., Goodrich, E. L., Bonaca, M. P., Mosenzon, O., Kato, E. T., Cahn, A., Furtado, R. H. M., Bhatt, D. L., Leiter, L. A., McGuire, D. K., Wilding, J. P. H., and Sabatine, M. S. (2019). SGLT2 inhibitors for primary and secondary prevention of cardiovascular and renal outcomes in type 2 diabetes: a systematic review and meta-analysis of cardiovascular outcome trials. *Lancet*.