Journal of Applied Statistics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/cjas20

Estimating effects of time-varying exposures on mortality risk

Trevor J. Thomson, X. Joan Hu & Bohdan Nosyk

To cite this article: Trevor J. Thomson, X. Joan Hu & Bohdan Nosyk (09 Feb 2024): Estimating effects of time-varying exposures on mortality risk, Journal of Applied Statistics, DOI: 10.1080/02664763.2024.2313459

To link to this article: https://doi.org/10.1080/02664763.2024.2313459

JOURNAL OF APPLIED STATISTICS
https://doi.org/10.1080/02664763.2024.2313459

Taylor & Francis
Taylor & Francis Group

Estimating effects of time-varying exposures on mortality risk

Trevor J. Thomson[a,b], X. Joan Hu[a] and Bohdan Nosyk[c,d]

[a]Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, British Columbia, Canada; [b]Fred Hutchinson Cancer Center, Biostatistics, Bioinformatics and Epidemiology Program, Seattle, WA, USA; [c]Faculty of Health Sciences, Simon Fraser University, Burnaby, British Columbia, Canada; [d]Centre for Health Evaluation & Outcome Sciences, St. Paul's Hospital, Vancouver, British Columbia, Canada

ABSTRACT

Administrative databases have become an increasingly popular data source for population-based health research. We explore how mortality risk is associated with some health service utilization process via linked administrative data. A generalized Cox regression model is proposed using a time-dependent stratification variable to summarize lifetime service utilization. Recognizing the service utilization over time as an internal covariate in the survival analysis, conventional likelihood methods are inapplicable. We present an estimating function based procedure for estimating model parameters, and provide a testing procedure for updating the stratification levels. The proposed approach is examined both asymptotically and numerically via simulation. We motivate and illustrate the proposed approach using an on-going program pertaining to opioid agonist treatment (OAT) management for individuals identified with opioid use disorders. Our analysis of the OAT data indicates that the OAT effect on mortality risk decreases in successive OAT attempts, in which two risk classes based on an individual's treatment episode number are established: one with 1–3 OAT episodes, and the other with 4+ OAT episodes.

ARTICLE HISTORY
Received 5 July 2023
Accepted 9 January 2024

KEYWORDS
Estimating equation; internal covariate; risk assessment; semiparametric regression; time-dependent stratification

1. Introduction

Administrative databases have become an increasingly popular data source to conduct population-based health research [13] due in part to its rich collection of service utilization records [15]. By deriving time-varying risk factors from these records, researchers have explored their association with a clinically meaningful event, such as an individual's mortality time [1,3,9]. This is typically done by specifying an extended Cox regression model, in which the time-varying exposures are specified as time-varying covariates. However, since individuals must be alive in order to have a health record, such time-varying covariates are internal [16]. The challenge brought on by internal covariates is that the conventional relationship between the hazard and survivor functions no longer holds, which prevents likelihood-based inference procedures to be adopted.

CONTACT X. Joan Hu joanh@stat.sfu.ca Department of Statistics and Actuarial Science, Simon Fraser University, 8888 University Drive, Burnaby V5A 1S6, BC, Canada

Supplemental data for this article can be accessed online at https://doi.org/10.1080/02664763.2024.2313459.

There are existing statistical approaches to overcome this challenge by re-framing the response or internal covariate processes. For example, one may use only part of the internal covariate history [26,32,33], or by introducing a multistate process in which levels of the internal covariate process and survival status define various states [2,8,10,11]. These approaches cannot yield the effect of the entire internal covariate process on mortality, and the multistate process demands modelling all of the state transitions. Alternatively, one may jointly model the event time and the internal covariate processes [23,31,34]. These approaches require modelling the underlying association of the two processes, and the implementation can be computationally expensive.

We propose a generalized Cox regression model with time-dependent stratification to formulate the effect of service utilization. Our modelling allows the data to report which strata are similar and naturally leads to dynamic grouping. This can aid in interpreting the analysis outcomes, and serve to improve the computational speed and the estimator's efficiency. We present an estimating function based procedure for model parameter estimation, and may be calculated by adapting the well-known R functions for survival analysis.

Motivated by the ongoing opioid epidemic, linked administrative health service records between 1996 and 2018 in British Columbia, Canada were provided to help improve the overall quality of care for individuals identified with an opioid use disorder [20]. We use this data to estimate the mortality risk of an individual given their current history of opioid agonist treatment (OAT) usage. As an individual must be alive to receive treatment, this variable is an internal covariate in the survival analysis. Previous studies have indicated that retention on an OAT can reduce the mortality risk of people with an opioid use disorder [17,25,27]. Revisiting such a study with administrative health records not only provides a real-world setting to generate scientific evidence on the clinical management of an opioid use disorder, but also control for additional time-varying exposures; a limitation with prior studies [25].

The rest of the paper is organized as follows. We introduce notation and the proposed modelling in Section 2. We detail the inference procedure in Section 3, and provide a straightforward testing procedure to update the levels of the stratification variable in a forward stepwise manner. We apply the proposed inference procedure to the linked administrative database that reports historical OAT dispensation records in Section 4. Based on the main findings from the data application, Section 5 summarizes the results from a simulation study. Some final remarks are given in Section 7.

2. Notation and modelling

Let T denote an individual's survival time since their first observed OAT dispensation, and $Z(t) \in \{0, 1\}$ denote an individual's OAT dispensation indicator at time $t \geq 0$, and $\mathcal{Z}(t) = \{Z(u) : 0 \leq u \leq t\}$. Additional covariates (baseline and time-varying) are denoted by $\boldsymbol{X}(t)$, with $\mathcal{X}(t) = \{\boldsymbol{X}(u) : 0 \leq u \leq t\}$ denoting its history up to time t, and let $\boldsymbol{W}(t) = (Z(t), \boldsymbol{X}(t)')'$ denote all covariates at time t. Consider a study with observations on T subject to a noninformative right-censoring time C. That is, the available information of T is the pair (T^*, δ), where $T^* = T \wedge C$ is the follow-up time of an individual, and $\delta = I(T \leq C)$ is the indicator for whether the survival time T is observed. Assume the study collects n independent and identically distributed realizations of $(T^*, \delta, \mathcal{Z}(T^*), \mathcal{X}(T^*))$. Our

objective can hence be understood as estimating the association between $\mathcal{Z}(\cdot)$ and T upon adjusting for $\mathcal{X}(\cdot)$ with right-censored observations of T.

We consider a generalized Cox regression model for the conditional hazard function of T given $\mathcal{Z}(t)$ and $\mathcal{X}(t)$ for $t > 0$:

$$\lambda(t; \mathcal{Z}(t), \mathcal{X}(t)) = \lambda_0(t; \mathcal{Z}(t)) \exp\{\boldsymbol{\theta}(\mathcal{Z}(t))' \boldsymbol{W}(t)\}, \tag{1}$$

where $\lambda_0(t; \mathcal{Z}(t))$ is an arbitrary baseline hazard function, and $\boldsymbol{\theta}(\mathcal{Z}(t))$ is a known function up to finite dimensional parameters, with the dimension of $\boldsymbol{\theta}(\mathcal{Z}(t))$ being the same as $\boldsymbol{W}(t)$. The model in (1) explicitly permits both the regression parameter and baseline hazard function to vary accordingly with an individual's current OAT dispensation history, in an attempt to adequately quantify the association between the dispensation process and mortality risk.

To specify this dependency we consider stratifying individuals into groups based on their dispensation history, which makes the stratification time-dependent. Let $g(\mathcal{Z}(t)) \in \{1, 2, \ldots, G\}$ denote a stratification variable that is fully determined by an individual's dispensation history up to time $t > 0$, where $G < \infty$ is known. That is, we let $\lambda_0(t; \mathcal{Z}(t)) = \lambda_{0g}(t)$ and $\boldsymbol{\theta}(\mathcal{Z}(t)) = \boldsymbol{\theta}_g = (\boldsymbol{\alpha}_g', \boldsymbol{\beta}')'$ when $g(\mathcal{Z}(t)) = g$, where $\boldsymbol{\theta}_g$ is a vector of unknown regression parameters. Here, $\boldsymbol{\alpha}_g$ is a q_A-dimensional vector of stratum-specific effects, and $\boldsymbol{\beta}$ is a q_B-dimensional vector of shared effects across strata. Without loss of generality, we partition the covariates as $\boldsymbol{W}(t) = (\boldsymbol{W}^A(t)', \boldsymbol{W}^B(t)')'$, where $\boldsymbol{W}^A(t)$ and $\boldsymbol{W}^B(t)$ have the same dimensions as $\boldsymbol{\alpha}_g$ and $\boldsymbol{\beta}$, respectively. The model in (1) then becomes

$$\lambda(t; \mathcal{Z}(t), \mathcal{X}(t)) = \lambda_{0g}(t) \exp\{\boldsymbol{\theta}_g' \boldsymbol{W}(t)\} \text{when } g(\mathcal{Z}(t)) = g, \text{ for } g = 1, \ldots, G, \tag{2}$$

which resembles an extended Cox regression model with a time-varying covariate $\boldsymbol{W}(t)$ and time-dependent strata [14].

A special case of (2) is if all of the baseline hazard functions are specified to be the same, so that the model reduces to

$$\lambda(t; \mathcal{Z}(t), \mathcal{X}(t)) = \lambda_0(t) \exp\{\boldsymbol{\theta}_g' \boldsymbol{W}(t)\} \text{when } g(\mathcal{Z}(t)) = g, \text{ for } g = 1, \ldots, G. \tag{3}$$

To assess if the model (3) fits the data better than (2), one could start by fitting the model (2), then plot the estimates of $\lambda_{01}(t), \cdots, \lambda_{0G}(t)$ over time, and assess if the model in (3) is appropriate. Alternatively, one could specify $\lambda_{0g}(t) = \lambda_{01}(t) \exp\{\gamma_g\}$ for $g = 2, \ldots, G$, and test whether $\gamma_g = 0$ for all $g = 2, \ldots, G$.

A special case of (3), and a further special case of (2), is obtained by not conducting any time-dependent stratification, in which we specify the model in (1) to be the extended Cox regression model [12]:

$$\lambda(t; \mathcal{Z}(t), \mathcal{X}(t)) = \lambda_0(t) \exp\{\boldsymbol{\theta}' \boldsymbol{W}(t)\}. \tag{4}$$

Clearly, we can recover (4) from (3) by testing $\boldsymbol{\alpha}_1 = \cdots = \boldsymbol{\alpha}_G$.

Since (2) is the most general of the three models presented, we take it as our primary model. The forthcoming estimation procedure is under (2), and we discuss how to slightly modify the estimation procedure to estimate parameters under (3) or (4). As the regression parameter captures the effect of the stratification variable in (3), the resulting estimates permit us to identify if the group effects between successive groups is significant. This leads us to dynamically update the stratification variable which not only improves the computing time to obtain the estimates, but also the statistical efficiency of the resulting estimator.

3. Estimation procedure

Let $N_i(t) = I(T_i \leq t)$, $Y_i(t) = I(T_i^* \geq t)$, and $\boldsymbol{\Theta} = (\boldsymbol{\alpha}_1', \ldots, \boldsymbol{\alpha}_G', \boldsymbol{\beta}')'$ be all the regression parameters in model (2), with $\boldsymbol{\Theta}_0$ denoting the true value of $\boldsymbol{\Theta}$.

3.1. Estimating regression parameters

Under (2), [22] showed that if all of the covariates in $\boldsymbol{W}(\cdot)$ are external, the partial score function of $\boldsymbol{\Theta}$ is $\boldsymbol{U}(\boldsymbol{\Theta}) = (U_1^A(\boldsymbol{\theta}_1)', \ldots, U_G^A(\boldsymbol{\theta}_G)', U^B(\boldsymbol{\Theta})')'$, where

$$U_g^A(\boldsymbol{\theta}_g) = \int_0^\infty \sum_{i:g(\mathcal{Z}_i(t))=g} Y_i(t) \left[\boldsymbol{W}_i^A(t) - \frac{E_g^A(t, \boldsymbol{\theta}_g)}{E_g(t, \boldsymbol{\theta}_g)} \right] dN_i(t), \quad g = 1, \ldots, G,$$

$$U^B(\boldsymbol{\Theta}) = \sum_{g=1}^G \int_0^\infty \sum_{i:g(\mathcal{Z}_i(t))=g} Y_i(t) \left[\boldsymbol{W}_i^B(t) - \frac{E_g^B(t, \boldsymbol{\theta}_g)}{E_g(t, \boldsymbol{\theta}_g)} \right] dN_i(t),$$

$$E_g(t, \boldsymbol{\theta}) = \sum_{j:g(\mathcal{Z}_j(t))=g} Y_j(t) \exp\{\boldsymbol{\theta}' \boldsymbol{W}_j(t)\},$$

$$E_g^C(t, \boldsymbol{\theta}) = \sum_{j:g(\mathcal{Z}_j(t))=g} Y_j(t) \exp\{\boldsymbol{\theta}' \boldsymbol{W}_j(t)\} \boldsymbol{W}_j^C(t), \quad C \in \{A, B\}.$$

One would then estimate $\boldsymbol{\Theta}$ with the solution to $\boldsymbol{U}(\boldsymbol{\Theta}) = \boldsymbol{0}$. However, as $Z(t)$ is an internal covariate, we can only view $\boldsymbol{U}(\boldsymbol{\Theta})$ as an estimating function for $\boldsymbol{\Theta}$. By applying modern empirical process theory [18], we show in the *Supplementary Material* that $\boldsymbol{U}(\boldsymbol{\Theta}_0)$ is centred at zero asymptotically under (2), and a reasonable estimator for $\boldsymbol{\Theta}$ is therefore the solution to $\boldsymbol{U}(\boldsymbol{\Theta}) = \boldsymbol{0}$. We also show in the *Supplementary Material* that $\hat{\boldsymbol{\Theta}}$ converges almost surely to $\boldsymbol{\Theta}_0$ under (2), and further establish the asymptotic distribution of $\sqrt{n}(\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}_0)$, in which the corresponding asymptotic variance of $\hat{\boldsymbol{\Theta}}$ can be consistently estimated with a Huber-like sandwich estimator $\widehat{\boldsymbol{AV}}(\hat{\boldsymbol{\Theta}}) = \hat{\boldsymbol{\Psi}}^{-1}(\hat{\boldsymbol{\Theta}}) \hat{\boldsymbol{\Phi}}(\hat{\boldsymbol{\Theta}}) \hat{\boldsymbol{\Psi}}^{-1}(\hat{\boldsymbol{\Theta}})$ with $\hat{\boldsymbol{\Psi}}(\boldsymbol{\Theta}) = -\frac{1}{n} \frac{\partial}{\partial \boldsymbol{\Theta}} \boldsymbol{U}(\boldsymbol{\Theta})$ and $\hat{\boldsymbol{\Phi}}(\boldsymbol{\Theta}) = \frac{1}{n} \sum_{i=1}^n \hat{\boldsymbol{\Omega}}_i(\boldsymbol{\Theta}) \hat{\boldsymbol{\Omega}}_i(\boldsymbol{\Theta})'$, where

$$\hat{\boldsymbol{\Omega}}_i(\boldsymbol{\Theta}) = (\hat{\Omega}_{i1}^A(\boldsymbol{\theta}_1)', \ldots, \hat{\Omega}_{iG}^A(\boldsymbol{\theta}_G)', \hat{\Omega}_i^B(\boldsymbol{\Theta})')',$$

$$\hat{\Omega}_{ig}^A(\boldsymbol{\theta}) = \int_0^\infty Y_i(t) I(g(\mathcal{Z}_i(t)) = g) \left[\boldsymbol{W}_i^A(t) - \frac{E_g^A(t, \boldsymbol{\theta})}{E_g(t, \boldsymbol{\theta})} \right] d\hat{M}_{ig}(t, \boldsymbol{\theta}),$$

$$\hat{\Omega}_i^B(\boldsymbol{\Theta}) = \sum_{g=1}^G \int_0^\infty Y_i(t) I(g(\mathcal{Z}_i(t)) = g) \left[\boldsymbol{W}_i^B(t) - \frac{E_g^B(t, \boldsymbol{\theta}_g)}{E_g(t, \boldsymbol{\theta}_g)} \right] d\hat{M}_{ig}(t, \boldsymbol{\theta}_g),$$

$$\hat{M}_{ig}(t, \boldsymbol{\theta}) = N_i(t) I(g(\mathcal{Z}_i(t)) = g) - \int_0^t Y_i(u) I(g(\mathcal{Z}_i(u)) = g) \exp\{\boldsymbol{\theta}' \boldsymbol{W}_i(u)\} d\hat{\Lambda}_{0g}(u),$$

and $\hat{\Lambda}_{0g}(\cdot)$ is a consistent estimator for $\Lambda_{0g}(\cdot)$. Observe that if all of the time-varying covariate(s) are indeed external covariate(s), we recognize $\hat{\boldsymbol{\Psi}}(\hat{\boldsymbol{\Theta}})$ as the observed information matrix under (2), so the corresponding variance estimator would simplify to $\widehat{\boldsymbol{AV}}(\hat{\boldsymbol{\Theta}}) = \hat{\boldsymbol{\Psi}}^{-1}(\hat{\boldsymbol{\Theta}})$.

We can estimate $\boldsymbol{\Theta} = \boldsymbol{\theta}$ under (4) by fixing $G \equiv 1$, so that $\boldsymbol{U}(\boldsymbol{\Theta}) = U^B(\boldsymbol{\Theta})$. We can also estimate the regression parameters in (3) by first including $\boldsymbol{W}^{A,(2)}(t), \ldots, \boldsymbol{W}^{A,(G)}(t)$ in $\boldsymbol{W}(t)$, where $\boldsymbol{W}^{A,(g)}(t) = I(g(\mathcal{Z}(t)) = g) \times \boldsymbol{W}^A(t)$, and then proceed as if the model is (4). Moreover, slight modifications to the arguments presented in the *Supplementary Material* establishes the large sample properties of the resulting estimators under (3) and (4).

3.2. Estimating baseline hazard functions

For fixed $g \in \{1, \ldots, G\}$, we view $d\Lambda_{0g}(t) = \lambda_{0g}(t)dt$ as a finite-dimensional parameter upon treating $\lambda_{0g}(\cdot)$ as a piece-wise constant function between uncensored survival times. With $\boldsymbol{\theta}_g$ fixed under (2), the following estimating equation is unbiased:

$$\sum_{i:g(\mathcal{Z}_i(t))=g} Y_i(t)[dN_i(t) - \exp\{\boldsymbol{\theta}_g' \boldsymbol{W}_i(t)\}\mathrm{d}\Lambda_{0g}(t)] = 0. \tag{5}$$

By solving for $\mathrm{d}\Lambda_{0g}(t)$, this promotes the estimator

$$\mathrm{d}\hat{\Lambda}_{0g}(t; \boldsymbol{\theta}_g) = \sum_{i:g(\mathcal{Z}_i(t))=g} \frac{Y_i(t)dN_i(t)}{\sum_{j:g(\mathcal{Z}_j(t))=g} Y_j(t) \exp\{\boldsymbol{\theta}_g' \boldsymbol{W}_j(t)\}}. \tag{6}$$

Here, we take the convention that $0/0 = 0$. By replacing the unknown $\boldsymbol{\theta}_g$ with its corresponding estimate, $\hat{\boldsymbol{\theta}}_g = (\hat{\boldsymbol{\alpha}}_g', \hat{\boldsymbol{\beta}}')'$, the baseline hazard function is estimated with a Breslow-like estimator $d\hat{\Lambda}_{0g}(t; \hat{\boldsymbol{\theta}}_g)$. Under the regularity conditions presented in the *Supplementary Material*, we can show that $d\hat{\Lambda}_{0g}(t; \hat{\boldsymbol{\theta}}_g)$ converges almost surely to $d\Lambda_{0g}(t)$. Furthermore, the weak convergence of $\mathrm{d}\hat{\Lambda}_{0g}(t; \hat{\boldsymbol{\theta}}_g)$ is also established, so that either $1 - \alpha^*$ pointwise confidence intervals or $1 - \alpha^*$ confidence bands for $\mathrm{d}\Lambda_{0g}(t)$ can be constructed.

If we instead consider the model in (3), the estimating equation (5) is slightly modified to

$$\sum_{g=1}^{G} \sum_{i:g(\mathcal{Z}_i(t))=g} Y_i(t)[dN_i(t) - \exp\{\boldsymbol{\theta}_g' \boldsymbol{W}_i(t)\}\mathrm{d}\Lambda_0(t)] = 0,$$

and $d\Lambda_0(t)$ is estimated with

$$\mathrm{d}\hat{\Lambda}_0(t; \boldsymbol{\theta}_g) = \sum_{g=1}^{G} \sum_{i:g(\mathcal{Z}_i(t))=g} \frac{Y_i(t)dN_i(t)}{\sum_{h=1}^{G} \sum_{j:g(\mathcal{Z}_j(t))=h} Y_j(t) \exp\{\boldsymbol{\theta}_h' \boldsymbol{W}_j(t)\}}. \tag{7}$$

The resulting estimator for $d\Lambda_0(t)$ under (4) then arises by fixing $\boldsymbol{\theta}_g = \boldsymbol{\beta}$ in (7).

3.3. Forward stepwise grouping based on Wald-type testing

In the case where $\boldsymbol{\alpha}_g$ fully captures the group effect and the levels of the stratification variable can be viewed as ordinal, one may question if the difference between successive groups are statistically significant. If the difference is not significant, we can simplify the model in (3) by merging groups $g-1$ and g together and re-estimate $\boldsymbol{\Theta}$ with the updated groups; otherwise, we keep these two groups separate from each other. Proceeding in this manner

would result in identifying $H \leq G$ data driven risk classes, which would provide a gain of efficiency by reducing the number of parameters to estimate.

To carry out this procedure, consider the following hypothesis test for a fixed $g \in \{2, \ldots, G\}$:

$$H_0 : \boldsymbol{\alpha}_g = \boldsymbol{\alpha}_{g-1} \quad \text{vs.} \quad H_a : \boldsymbol{\alpha}_g \neq \boldsymbol{\alpha}_{g-1}. \tag{8}$$

Based on the asymptotic normality of $\hat{\boldsymbol{\Theta}}$, we can construct a Wald test statistic

$$J_g = (\hat{\boldsymbol{\alpha}}_g - \hat{\boldsymbol{\alpha}}_{g-1})' \Big[\text{Var}(\hat{\boldsymbol{\alpha}}_g - \hat{\boldsymbol{\alpha}}_{g-1}) \Big]^{-1} (\hat{\boldsymbol{\alpha}}_g - \hat{\boldsymbol{\alpha}}_{g-1}). \tag{9}$$

With the variance of $\hat{\boldsymbol{\Theta}}$ estimated with $\widehat{\boldsymbol{AV}}(\hat{\boldsymbol{\Theta}})$, we can therefore estimate $\text{Var}(\hat{\boldsymbol{\alpha}}_g - \hat{\boldsymbol{\alpha}}_{g-1})$ with $\boldsymbol{C}_g \widehat{\boldsymbol{AV}}(\hat{\boldsymbol{\Theta}}) \boldsymbol{C}_g'$, where $\boldsymbol{C}_g$ is a constant such that $\boldsymbol{C}_g \boldsymbol{\Theta} = \boldsymbol{\alpha}_g - \boldsymbol{\alpha}_{g-1}$. Under H_0 in (8), $J_g \dot{\sim} F_{q_A}(\cdot)$, where $F_{q_A}(\cdot)$ denotes the χ^2-distribution function with q_A degrees of freedom. We hence reject H_0 if $J_g > F_{q_A}^{-1}(1 - \alpha^*)$, where α^* is the type I error rate. As the test proceeds in a forward stepwise manner, we refer to this test as the *forward stepwise Wald test*.

4. Analysis of provincial OAT dispensation records

We applied the proposed inference procedure to the provincial health administrative database. Specifically, individuals identified with an opioid use disorder between 01/01/1996 and 10/01/2018 in British Columbia, Canada, health records in the form of drug dispensations (British Columbia Ministry of Health [6]), hospital and emergency department admissions (British Columbia Ministry of Health [4]), physician billing records (British Columbia Ministry of Health [5]), incarceration records (Ministry of Public Safety and Solicitor General [19]), and deaths (British Columbia Vital Statistics Agency [7]) were provided. Only individuals with at least one OAT dispensation and at least 18 years old on their first recorded OAT dispensation date were included in the analysis. In total, there are $n = 54{,}739$ individuals included in our study.

We classified individuals to receive an OAT if they had a pharmaceutical health record indicating they were dispensed OAT. Since dispensation records were unavailable for inpatient services, we follow [21], and assumed that an OAT was provided to an individual during their (entire) hospitalization if they either had an OAT dispensation record within five days (i) before their admission date, or (ii) after their discharge date. Risk factors included in our analysis were based on a preliminary analysis conducted by [21]. The time-independent covariates we included are *sex* (male vs. female), *birth generation* (indicators for birth year: 1901–1945 vs. 1946–1964 vs. 1965–1980 vs. 1981+), *health authority* (indicators of residence region: Fraser Health vs. Interior vs. Vancouver Coastal vs. Vancouver Island vs. Northern), and *year category* (category corresponding to first observed OAT dispensation date: 1996–2000 vs. 2001–2006 vs. 2007–2012 vs. 2013–2018). The time-varying covariates we included are *OAT dispensation indicator* (by time t), *alcohol or other substance use disorders* (by time t), *mental illness or chronic pain* (by time t), *HCV or HIV/AIDS* (hepatitis C virus or human immunodeficiency virus or acquired immunodeficiency syndrome by time t), *indicator of ever receiving a sedative*, (use of a sedative by time t), and *ever on PharmaCare Plans C or G* (indicator of poor socio-economic status by time t). Although

Table 1. Estimates of regression coefficients under the Cox regression model (4).

Covariate names	Estimate	S.E.	*P*-value*
OAT Status	**−1.0182**	0.0265	< 0.0001
Sex (vs. *Female*)	–	–	–
Male	**0.2239**	0.0261	< 0.0001
Birth Generation (vs *Greatest & Silent Generations*)	–	–	–
Baby Boomers	−1.3215	0.0443	< 0.0001
Generation X	**−2**.1372	0.0478	< 0.0001
Millennials & Generation Z	**−2**.2044	0.0591	< 0.0001
Heath Authority (vs *Fraser Health*)	–	–	–
Interior	**0.1455**	0.0387	0.0002
Vancouver Coastal	**0.1041**	0.0302	0.0006
Vancouver Island	0.0674	0.0362	0.0624
Northern	0.0015	0.0677	0.9822
Year Category (vs. *1996–2000*)	–	–	–
2001–2006	**0.0789**	0.0330	0.0166
2007–2012	**0.1509**	0.0391	0.0001
2013–2018	**0.4733**	0.0492	< 0.0001
Alcohol or Other Substance Use Disorders	**0.4492**	0.0424	< 0.0001
Ill Mental Health or Chronic pain	**−0.1898**	0.0431	< 0.0001
Hepatitis C Virus or HIV/AIDS	**1.1748**	0.0272	< 0.0001
Ever Received a Sedative	**0.4756**	0.0332	< 0.0001
Ever on PharmaCare Plans C or G	0.0290	0.0299	0.3329
Incarceration Status	**−1.6318**	0.1782	< 0.0001
Number of Incarcerations	**0.0061**	0.0028	0.0305

∗: Under the hypothesis of a null effect. Note: The reported standard-error (S.E.) estimates of $\hat{\boldsymbol{\Theta}}$ correspond to the square-root of the diagonal elements of $\widehat{AV}(\hat{\boldsymbol{\Theta}})$. Estimates that are **bold-faced** are statistically significant with the type 1 error rate set at $\alpha^* = 5\%$.

the variable *health authority* is in principle a time-varying covariate, the data showed this variable to be rather stable, and is therefore treated as a time-independent variable. There were also few changes in *sex* over time, many of which were attributed to data entry errors.

We began our analysis by fitting an extended Cox regression model (4) to the observed data. The parameter estimates under *time since first observed OAT dispensation time scale* are displayed in Table 1, where we can see that the effect of *OAT dispensation indicator* is negative and statistically significant, which corroborates with prior studies. We remark that the first *observed* OAT dispensation may not necessarily corresponding to the true actual first OAT dispensation for an individual, since the dispensation records only captures records within British Columbia during the data extraction window. The implications of this is that time zero may be informative. We additionally consider *age* as the time scale, where the results are presented in the *Supplementary Material* in Table S1.

We proceeded to fit a stratified Cox regression model (2), in which individuals were stratified according to their OAT episode number at time t, where an OAT episode at time t is the number of long term *not dispensed OAT* to *dispensed OAT* transitions by time t, as illustrated in Figure 1. We specified $G = 9$ levels for the stratification variable: (i) 1 OAT episode; (ii) 2–3 OAT episodes; (iii) 4–5 OAT episodes; (iv) 6–7 OAT episodes; (v) 8–10 OAT episodes; (vi) 11–15 OAT episodes; (vii) 16–20 OAT episodes; (viii) 21–30 OAT episodes; and (ix) 30+ OAT episodes. The G levels were selected based on a combination of summary statistics for the number of OAT episodes individuals experienced by their end of follow-up date, as well as expert opinion. The estimates of $\boldsymbol{\Theta}$ with $\boldsymbol{\theta}_g = \boldsymbol{\alpha}_g$ when $g(\mathcal{Z}(t)) = g$, are presented in Figure 2. We see the *OAT dispensation indicator*, the *birth*

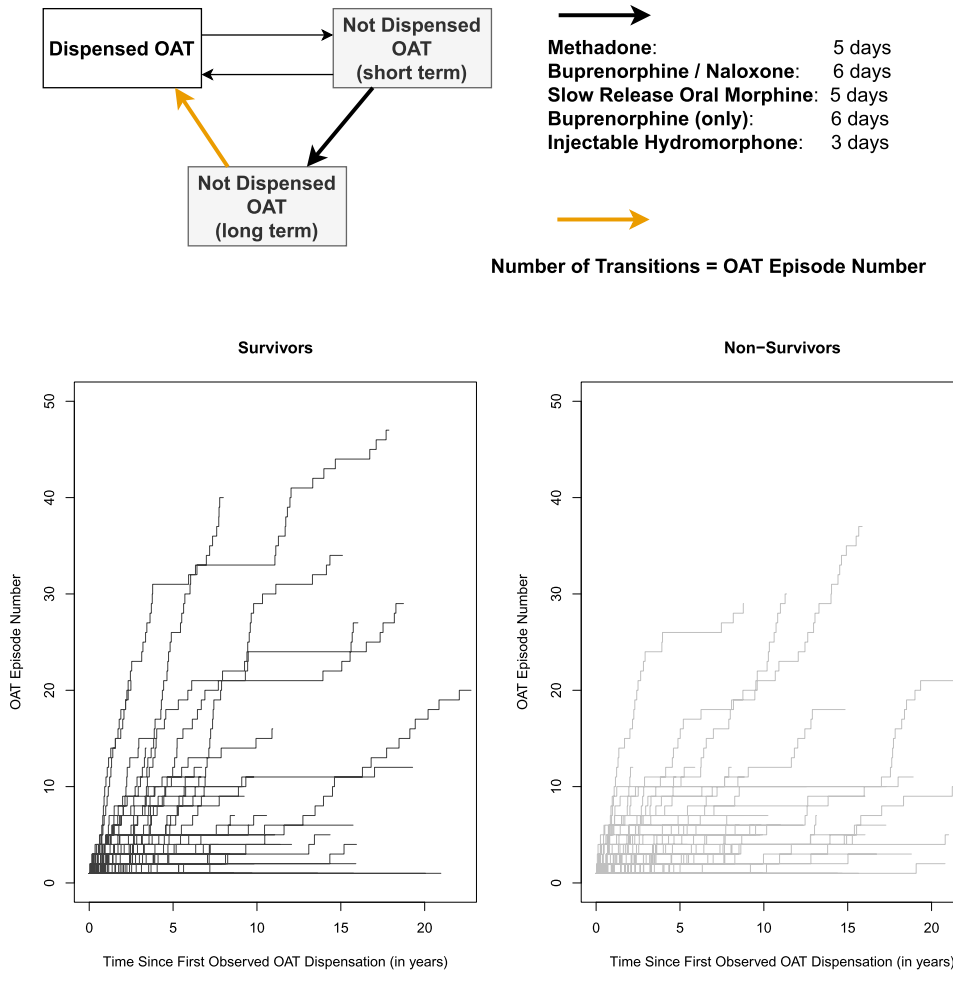

Figure 1. Top: A multistate representation of the OAT dispensation process. The OAT episode of an individual at time t is the number of (long-term) *not dispensed OAT* to *dispensed OAT* transitions they experience by time t, with individuals initializing in OAT episode 1. **Bottom**: A plot of 100 randomly selected survivors and non-survivors and their OAT episode numbers over time.

generation indicators, and the *ever on PharmaCare plans C or G* indicator appear to have varying effects across strata, whereas the other effects are constant. We also present the LOESS-smoothed estimates of $\lambda_{0g}(\cdot)$ in Figure 3, in which we see the estimates overlap with one another, aside from the groups corresponding to "large" OAT episode groups, and motivates us to consider fitting (3). The corresponding estimates with *age* as the time scale are presented in Figures S1 and S2.

We proceeded to fit model (3) with the same stratification variable, and $\boldsymbol{\theta}_g = \boldsymbol{\alpha}_g$ when $g(\mathcal{Z}(t)) = g$, which produced similar results to Figures 2 and S1. We updated our modelling by specifying the constant effects to be $\boldsymbol{\beta}$, which will serve to reduce the computational intensity by estimating fewer parameters, as well as improve the efficiency of the estimates. The parameter estimates upon specifying $\boldsymbol{\theta}_g = (\boldsymbol{\alpha}_g', \boldsymbol{\beta}')'$ when $g(\mathcal{Z}(t)) = g$ are illustrated in Figure 4 and Table 2. We can see that the estimates that vary across strata are

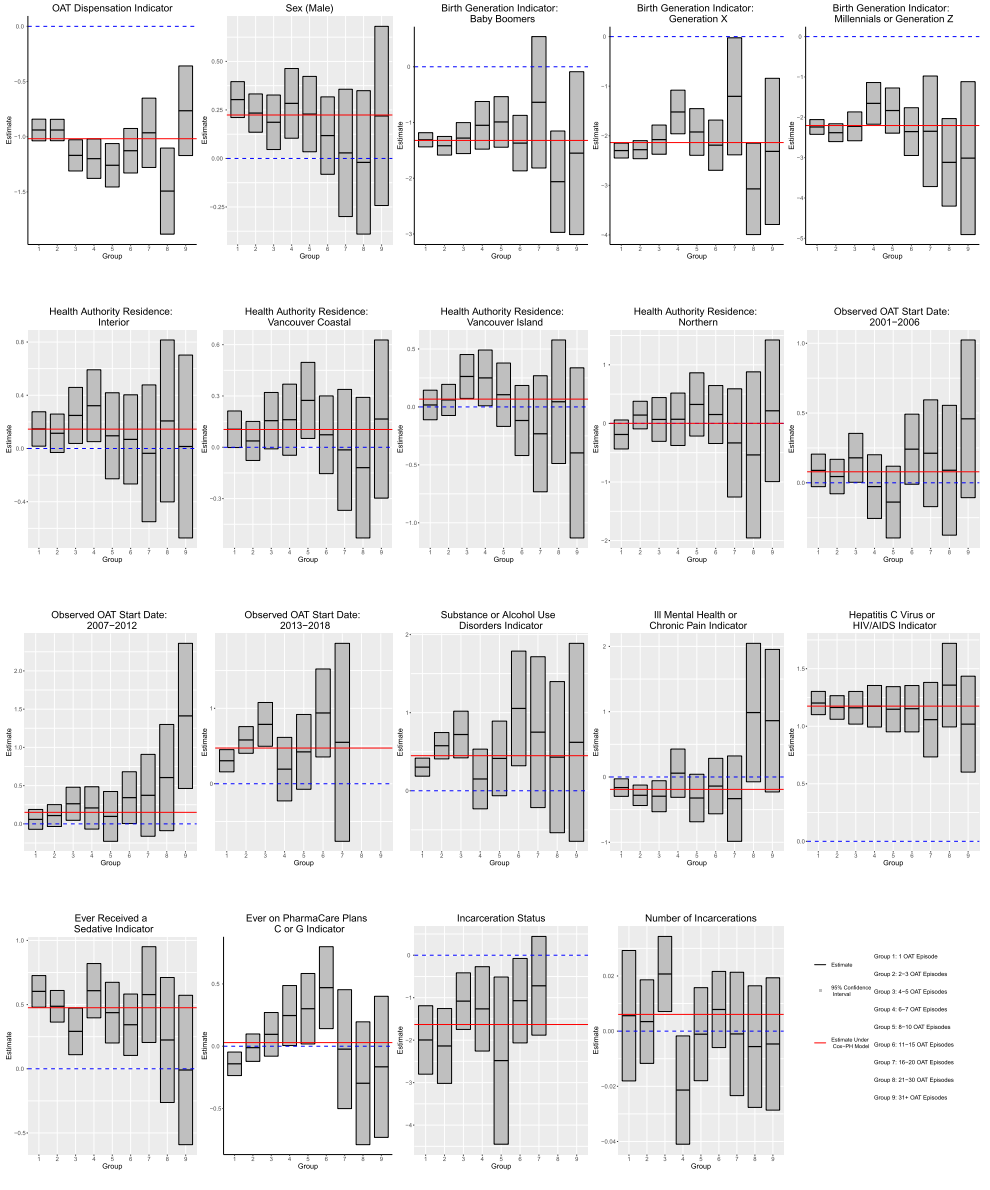

Figure 2. Estimates of regression coefficients under the stratified Cox regression model (2), where the time scale is *time since first observed OAT dispensation*, and $\boldsymbol{\theta}_g = \boldsymbol{\alpha}_g$. Variables with a gray background appear to have a constant effect across strata. As a reference, we illustrate the estimated effect under the Cox model (4) with a red line.

quite similar to their corresponding estimates in Figure 2, and the estimates in Table 2 are similar to their corresponding estimates shown in Table 1. The corresponding estimates with *age* as the time scale are presented in Figure S3 and Table S2.

We proceeded to conduct the forward stepwise Wald test in Section 3.3 to update the stratification variable, where the results are shown in Tables S4 and S5. For each test, we

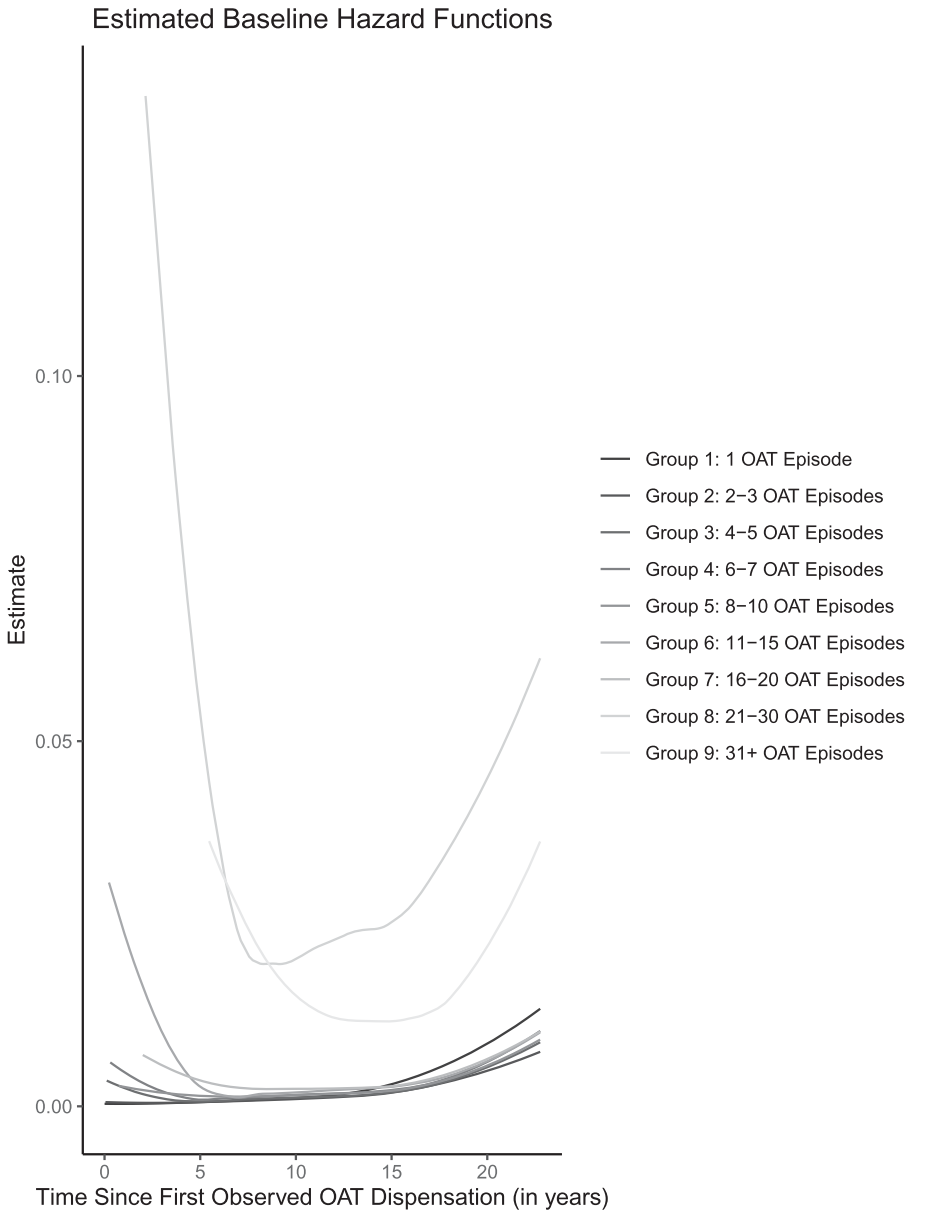

Figure 3. Smoothed estimates of $\lambda_{0g}(\cdot)$ under (2), where we stratify by the OAT episode number at time t.

displayed the estimates of both $\boldsymbol{\alpha}_{g-1}$ and $\boldsymbol{\alpha}_g$, the test statistic J_g in (9), and the resulting p-value. By applying a Bonferonni correction for the multiple testing, we specified the type I error rate to be $\alpha^* = 0.05/8$. The results of the test under both *time since first observed OAT dispensation* and *age* time scales reveals that the stratification variable should be updated to the following $H = 2$ levels: (i) 1-3 OAT episodes; and (ii) 4+ OAT episodes. The results with the updated stratification are illustrated in Figure 5 and Table 3, in which we see the *OAT*

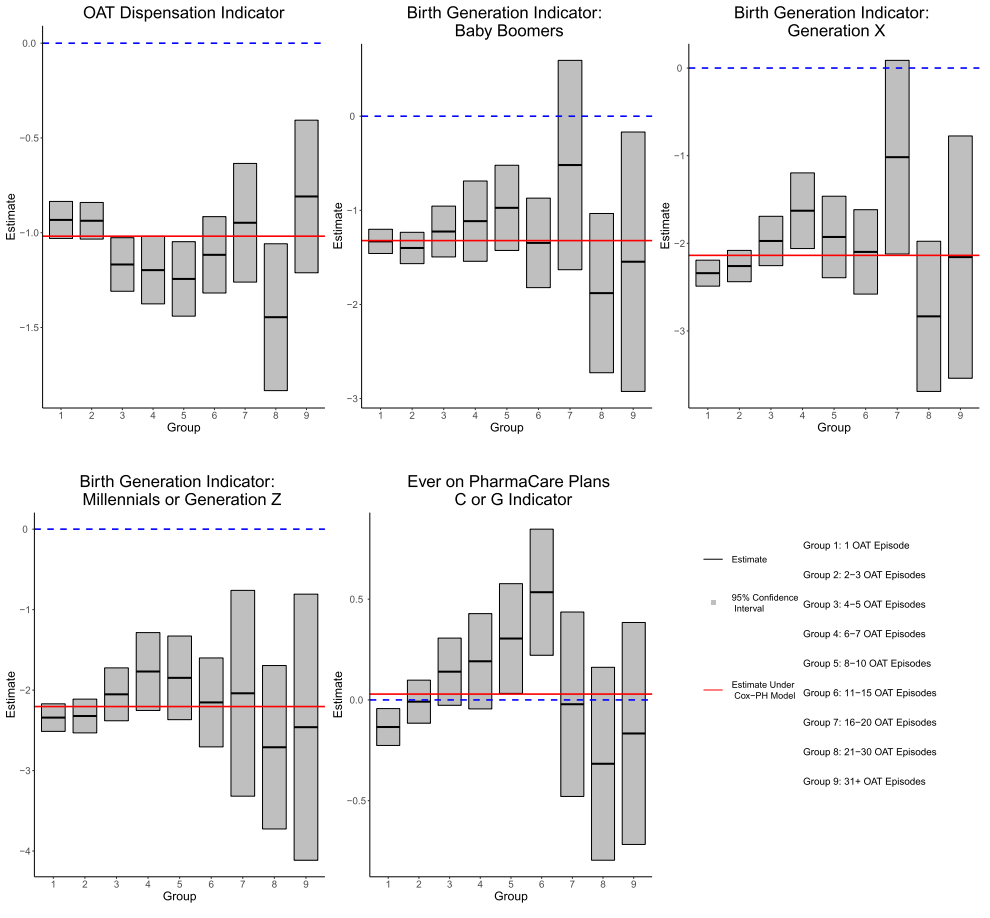

Figure 4. Estimates of regression coefficients under the stratified Cox regression model (3), where the time scale is *time since first observed OAT dispensation*. Variables with effect $\boldsymbol{\alpha}_g$ are illustrated below.

dispensation indicator effect has a higher protective effect against mortality for individuals with more OAT episodes, whereas the opposite is true for the *birth generation indicator* and *ever on PharmaCare plans C or G* effects. The corresponding estimates with *age* as the time scale are presented in Figure S4 and Table S3.

5. Simulation study

We conducted three simulation studies to jointly examine the finite-sample performance of the proposed estimator and assess the performance of the forward stepwise Wald test. Specifically, we generated data based on the Cox regression model in the first simulation study, generated data based on the stratified Cox regression model with a shared baseline hazard function in the second simulation study, and assessed the robustness of the forward stepwise Wald test against model misspecification in the third simulation study. We summarized our simulation results into tables presented as *Supplementary Material*.

Table 2. Estimates of regression coefficients under the stratified Cox regression model (3), where the time scale is *time since first observed OAT dispensation*.

Covariate names	Estimate	S.E.	*P*-Value*
Sex (vs. *Female*)	–	–	–
Male	**0.2340**	0.0260	< 0.0001
Heath Authority (vs *Fraser Health*)	–	–	–
Interior	**0.1535**	0.0385	0.0001
Vancouver Coastal	**0.1079**	0.0301	0.0003
Vancouver Island	**0.0723**	0.0366	0.0484
Northern	0.0114	0.0675	0.8653
Year Category (vs. *1996–2000*)	–	–	–
2001–2006	**0.0791**	0.0338	0.0191
2007–2012	**0.1430**	0.0387	0.0002
2013–2018	**0.4619**	0.0498	< 0.0001
Alcohol or Other Substance Use Disorders	**0.4329**	0.0447	< 0.0001
Ill Mental Health or Chronic pain	**−0.1887**	0.0434	< 0.0001
Hepatitis C Virus or HIV/AIDS	**1.1704**	0.0273	< 0.0001
Ever Received a Sedative	**0.4791**	0.0335	< 0.0001
Incarceration Status	**−1.6286**	0.1726	< 0.0001
Number of Incarcerations	0.0014	0.0030	0.6492

∗: Under the hypothesis of a null effect. Note: Variables with effect $\boldsymbol{\beta}$ are tabulated below.

Table 3. Estimates of regression coefficients under the stratified Cox regression model (3) following the forward stepwise Wald test, where the time scale is *time since first observed OAT dispensation*, and $\boldsymbol{\theta}_g = (\boldsymbol{\alpha}_g', \boldsymbol{\beta}')'$.

Covariate names	Estimate	S.E.	*P*-Value*
Sex (vs. *Female*)	–	–	–
Male	**0.2327**	0.0262	< 0.0001
Heath Authority (vs *Fraser Health*)	–	–	–
Interior	**0.1517**	0.0387	0.0001
Vancouver Coastal	**0.1062**	0.0302	0.0004
Vancouver Island	0.0706	0.0362	0.0514
Northern	0.0062	0.0676	0.9273
Year Category (vs. *1996–2000*)	–	–	–
2001–2006	**0.0819**	0.0329	0.0129
2007–2012	**0.1488**	0.0391	0.0001
2013–2018	**0.4706**	0.0492	< 0.0001
Alcohol or Other Substance Use Disorders	**0.4457**	0.0423	< 0.0001
Ill Mental Health or Chronic pain	**−0**.1904	0.0430	< 0.0001
Hepatitis C Virus or HIV/AIDS	**1.1725**	0.0273	< 0.0001
Ever Received a Sedative	**0.4768**	0.0333	< 0.0001
Incarceration Status	**−1**.6244	0.1793	< 0.0001
Number of Incarcerations	0.0023	0.0029	0.4324

∗: Under the hypothesis of a null effect. Note: Variables with effect $\boldsymbol{\beta}$ are tabulated below.

5.1. Data generation

We generated right-censored observations of an event time, and an alternating binary process that affects the event time's hazard. We specified $n = 10{,}000$ independent units, so that the number of units roughly matches the sample size in our data application. For each study unit, we generated observations as follows:

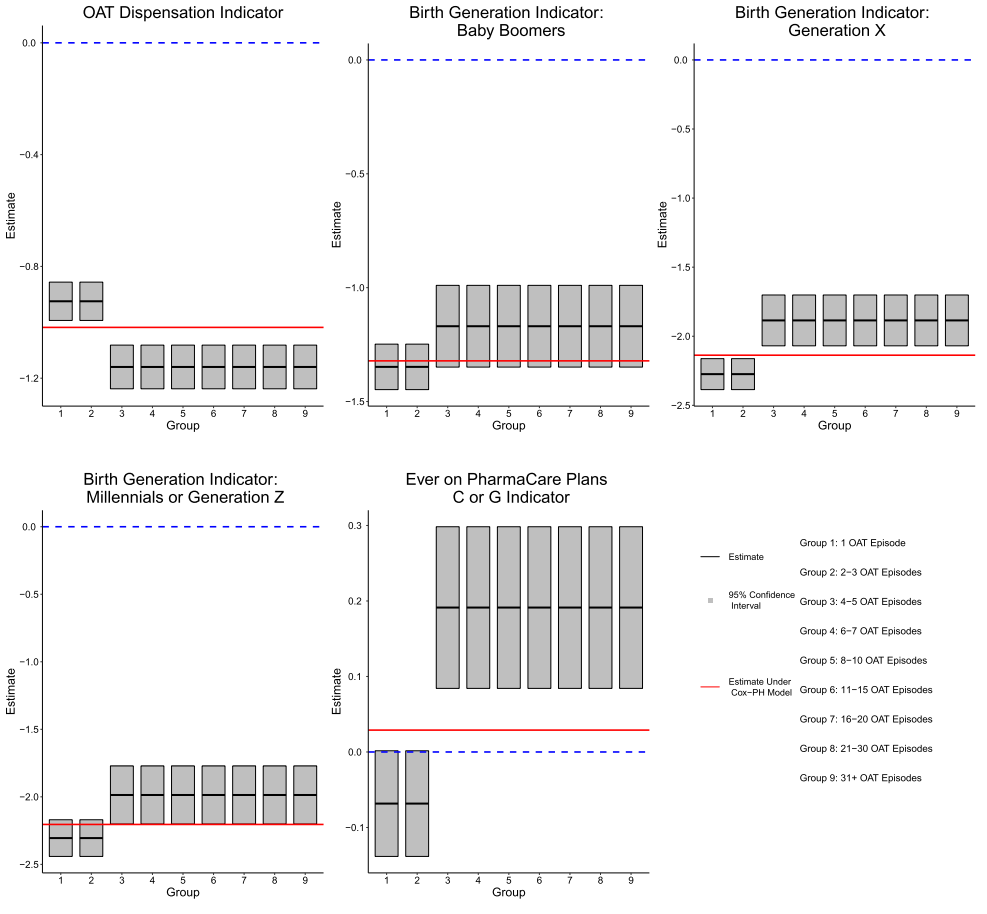

Figure 5. Estimates of regression coefficients under the stratified Cox regression model (3) following the forward stepwise Wald test, where the time scale is *time since first observed OAT dispensation*, and $\boldsymbol{\theta}_g = (\boldsymbol{\alpha}_g', \boldsymbol{\beta}')'$. Variables with effect $\boldsymbol{\alpha}_g$ are illustrated below.

(i) Generate two baseline covariates, X_1 and X_2, where $X_1 \sim \text{Uniform}(0, 1)$, and $X_2 \sim \text{Bernoulli}(0.5)$.

(ii) Generate a time-varying alternating binary indicator, $Z(t)$. To be consistent with our data application, we specified $Z(0) \equiv 1$, $g(\mathcal{Z}(0)) \equiv 1$, and $g(\mathcal{Z}(t))$ is determined by the number of zero-to-one changes up to time t. We specified $G = 10$, and keep units with more than 10 zero-to-one changes in group G. To generate $Z(t)$, we simulated the time an individual changes their binary indicator status from the exponential distribution, where the rates $\rho_0(\mathcal{Z}(t)) = \rho_{0g}$ and $\rho_1(\mathcal{Z}(t)) = \rho_{1g}$ for the time $Z(t)$ transitions to 0 and 1, respectively, when $g(\mathcal{Z}(t)) = g$. Here, we specified

$$\rho_{0g} = \begin{cases} 20 & \text{if} g \in \{1,2,3\} \\ 30 & \text{if} g \in \{4,5,6\} \\ 50 & \text{if} g = 7 \\ 10 & \text{if} g \in \{8,9,10\} \end{cases} \quad \text{and} \quad \rho_{1g} = \begin{cases} 10 & \text{if} g \in \{1,2,3\} \\ 20 & \text{if} g \in \{4,5,6\} \\ 50 & \text{if} g = 7 \\ 5 & \text{if} g \in \{8,9,10\} \end{cases}.$$

We remark that a 'large' sample size is needed to ensure that enough information is available to estimate $\boldsymbol{\theta}_g$ within a simulated dataset, especially for large values of g.

(iii) With $\boldsymbol{W}(t) = (Z(t), X_1, X_2)'$ as the vector of covariates at time t, we discretized the time interval for which $Z(t)$ is constant into subintervals of length $\Delta t = 0.0001$, and simulated the event occurrence at time t as a Bernoulli random variable with success probability $\lambda^{\dagger}(t; \mathcal{Z}(t), X_1, X_2) \times \Delta t$, where

$$\lambda^{\dagger}(t; \mathcal{Z}(t), X_1, X_2) = \lambda_0 f(\boldsymbol{\theta}_g' \boldsymbol{W}(t)),$$

with $f(x) = e^x$ or $f(x) = (1 + x)I(x \geq 0) + (1 - x)I(x < 0)$, and the specification of $f(\cdot)$ depends on the particular simulation outcome presented in Sections 5.2, 5.3, and 5.4. This procedure continues until we observe a success.

(iv) Generate (non-informative) censoring times from the exponential distribution with rate $\lambda_C \in \{0.5, 1.5\}$ to produce right-censored event times.

Overall, this data generation procedure produces the following independent observations

$$< /p >< p > \{(T_i^*, \delta_i, \mathcal{Z}_i(T_i^*), X_{i1}, X_{i2}) : i = 1, \ldots, n\},$$

with $T_i^* = T_i \wedge C_i$, and $\delta_i = I(T_i \leq C_i)$. With a simulated dataset, we proceeded to estimate $\boldsymbol{\Theta}$ and conduct the forward stepwise Wald test. We replicated the data generation and inference procedure 150 times.

5.2. Simulation outcome: reduction to the Cox regression model

We first conducted a simulation study where we generated event times under the Cox regression model with the baseline hazard function specified as a constant over time. That is, we generated event times under the hazard model

$$\lambda^{\dagger}(t; \mathcal{Z}(t), X_1, X_2) = \lambda_0 \exp\{\boldsymbol{\theta}' \boldsymbol{W}(t)\},$$

with $\lambda_0 = 1.75$ and $\boldsymbol{\theta} = (-1, 0, -0.5)'$. Based on the two specifications of λ_C, this resulted in (on average) approximately 29% and 52% of event times being right-censored. We considered two approaches to estimate $\boldsymbol{\theta}$: (a) solve $\boldsymbol{U}(\boldsymbol{\theta}) = \boldsymbol{0}$ in which we fix $G \equiv 1$, and (ii) use the `coxph` function in the `survival` R package [29], where the results are presented in Table S6. As expected, we see that the estimates of $\boldsymbol{\theta}$ are equivalent across the two procedures, and $\hat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{AV}}(\hat{\boldsymbol{\theta}})$ are consistent estimators for $\boldsymbol{\theta}$ and $\boldsymbol{AV}(\boldsymbol{\theta})$, respectively. Note that the estimated standard errors appear to be similar under the two approaches. This is likely due to fitting the appropriate data generating model to the data, which is resulting in $\boldsymbol{\Psi}^{-1}(\boldsymbol{\theta}) \approx \boldsymbol{\Phi}(\boldsymbol{\theta})$.

We then fit (3) to the simulated data, where we present the results in Table S7. We observed the standard error estimates to be larger relative to the standard error estimates under the Cox regression model. This is due to all n units estimating $\boldsymbol{\theta}$ in the Cox regression model, whereas only units in group g contribute to the estimation of $\boldsymbol{\theta}_g$. We proceeded to conduct the forward stepwise Wald test from Section 3.3, and assess its performance in correctly recovering the Cox regression model. We present matrices in Table S8, where the (g, g') element is the proportion groups g and g' are classified to the same class, with

$\alpha^* = 0.05$. The matrices inform us that there is approximately a 95% chance of group $g-1$ (correctly) being merged together with group g. Therefore, we would anticipate that the Cox regression model should be recovered approximately 55% of the time. In order to correctly recover the Cox regression model 95% of the time, we propose a Bonferonni correction and specify $\alpha^* = 0.05/9$. As shown in Table S9, we see that the test can adequately recover the Cox regression model.

5.3. Simulation outcome: correctly identifying the number of risk classes in the stratified Cox model

We conducted a simulation study where we generated event times under the stratified Cox regression model with a constant baseline hazard function. That is, we generated event times under the following hazard model

$$\lambda^{\dagger}(t; \mathcal{Z}(t), X_1, X_2) = \lambda_0 \exp\{\boldsymbol{\theta}_g' \boldsymbol{W}(t)\},$$

with $\lambda_0 = 1.75$, and $\boldsymbol{\theta}_g = \begin{cases} (-0.5, -2, -2)' & \text{if} g \in \{1, 2, 3\} \\ (-1, 0, -0.5)' & \text{if} g \in \{4, 5, 6\} \\ (-2, 2, 1.5)' & \text{if} g \in \{7, 8, 9, 10\} \end{cases}$. Based on the two specifications of λ_C, this resulted in (on average) approximately 38% and 73% of event times being right-censored.

We started out by naively fitting the Cox regression model in (4), where our results are summarized in Table S10. We note that the resulting estimates resemble a weighted average of the $\boldsymbol{\theta}_g$ parameters across the G groups. Furthermore, our standard error estimates were generally larger relative to the standard error estimates reported by the `coxph` function, as expected.

We then proceeded to fit the true data generating model in (2) to the simulated data. The results are summarized in Table S11. We observed that the estimates $\hat{\boldsymbol{\Theta}}$ and $\widehat{\boldsymbol{AV}}(\hat{\boldsymbol{\Theta}})$ appear to consistently estimate $\boldsymbol{\Theta}$ and $\boldsymbol{AV}(\boldsymbol{\Theta})$, respectively. Similar to Section 5.2, we generally saw the standard error estimates for our first two parameter settings to be smaller than the standard error estimates for the last two parameter settings, which is attributed to the smaller censoring rate.

We then conducted the forward stepwise Wald test to assess its performance in correctly recovering the three data-generating treatment classes. Our results illustrated in Table S12 informs us that there is approximately a 95% chance that group $g-1$ is merged together with group g if they belong to the same class. In fact, groups $g-1$ and g were never combined together if they do not belong to the same risk class. By applying a Bonferonni correction, our test performed adequately in recovering the three risk classes.

5.4. Simulation outcome: robustness to model misspecification

We conducted a simulation study where we generated event times under a misspecified stratified Cox regression model. Specifically, we generated event times under the hazard model

$$\lambda^{\dagger}(t; \mathcal{Z}(t), X_1, X_2) = \lambda_0 f(\boldsymbol{\theta}_g' \boldsymbol{W}(t)),$$

where $\lambda_0 = 1.75$, and $f(x) = (1+x)I(x \geq 0) + (1-x)I(x < 0)$ for $g = 1, \ldots, 10$. Upon specifying $\lambda_C = 0.5$ and $\lambda_C = 1.5$, this resulted in (on average) approximately 29% and 62% of individuals having right-censored death times, respectively. The purpose of this simulation study is to assess if the forward stepwise Wald test is robust to model misspecification.

By estimating the model parameters under the stratified Cox regression model (2), Tables S13 and S14 shows that the forward stepwise Wald test is robust to model misspecification upon applying a Bonferonni correction.

6. On sample size determination

Our simulation study demonstrates that we can adequately recover the true risk classes with our specification of $\boldsymbol{\Theta}$, where the type I error rate is $\alpha^* = 0.05/9$, and $n = 10{,}000$. We anticipate a larger sample size is required to correctly identify the correct risk classes if we instead specify difference between stratum-specific effects across successive groups to be "small". We now consider the minimum sample size needed for the forward stepwise Wald test to achieve a certain power, $1-\beta^*$ given α^*, and effect difference $\boldsymbol{\alpha}_g - \boldsymbol{\alpha}_{g-1}$. In other words, consider the following simple hypothesis for a fixed $g \in \{2, \ldots, G\}$:

$$H_0 : \boldsymbol{\alpha}_g - \boldsymbol{\alpha}_{g-1} = \mathbf{0} \quad \text{vs.} \quad H_a : \boldsymbol{\alpha}_g - \boldsymbol{\alpha}_{g-1} = \boldsymbol{\gamma}_g \neq 0. \tag{10}$$

Here, $\boldsymbol{\gamma}_g$ is the smallest value in which we view the difference in effects to be meaningful. In order for the forward stepwise Wald test to conduct the hypothesis test in (10) adequately, we want to reject H_0 with probability $1-\beta^*$ when H_a is true, and reject H_0 with probability α^* when H_0 is true. Here, β^* is referred to as the type II error rate, and $1-\beta^*$ is referred to as the power of the test. Recall in Section 3.3 that under H_0 in (10), $J_g \dot{\sim} F_{q_A}(\cdot)$, where $F_{q_A}(\cdot)$ denotes the χ^2-distribution function with q_A degrees of freedom. However, under H_a in (10), $J_g \dot{\sim} F_{q_A,\nu_g}(\cdot)$, where $F_{q_A,\nu_g}(\cdot)$ denotes the non-central χ^2-distribution with q_A degrees of freedom with non-centrality parameter ν_g, with

$$\nu_g = (\boldsymbol{\alpha}_g - \boldsymbol{\alpha}_{g-1})' \left[\frac{\text{Var}(\boldsymbol{\alpha}_g - \boldsymbol{\alpha}_{g-1})}{n} \right]^{-1} (\boldsymbol{\alpha}_g - \boldsymbol{\alpha}_{g-1}). \tag{11}$$

Then for α^* and β^* given, we want to find ν_g such that

$$F_{q_A,\nu_g}(F_{q_A}^{-1}(1-\alpha^*)) = \beta^*.$$

Hence, we can proceed to solve for n in (11) as

$$n_g = \nu_g \left[(\boldsymbol{\alpha}_g - \boldsymbol{\alpha}_{g-1})' \text{Var}(\boldsymbol{\alpha}_g - \boldsymbol{\alpha}_{g-1})^{-1} (\boldsymbol{\alpha}_g - \boldsymbol{\alpha}_{g-1}) \right]^{-1}.$$

But since we considered the test (10) for a fixed g, we iterate over g to obtain $n_2, \ldots, n_G$, and conclude that the sample size needed to correctly recover the correct group structure

with type I error α^* and type II error β^* is

$$n = \max\{n_g : g = 2, \ldots, G\}.$$

7. Conclusion

We proposed a generalized Cox regression model, under which we conducted time-dependent stratification, where the strata are defined in terms of the history of an internal time-varying covariate process. We adopted an estimating-equation based inference procedure that bypasses interpretation challenges brought on by constructing the likelihood function. Large sample properties of the proposed estimators were established, and our simulation study shows that we are able to consistently estimate both the regression parameters and the estimate's standard error. To determine if the effects between successive groups are significant, we proposed a Wald test that sequentially tests if groups should be merged. Upon applying a Bonferonni correction, we showed through a simulation study that this test can correctly recover the true grouping structure in a satisfactory manner, and is robust to model misspecification. We applied the proposed methodology to a provincial health administrative database pertaining to individuals with at least one OAT dispensation record, in which two risk classes based on an individual's history of OAT use were identified. We summarized an individual's history of the OAT dispensation process with their OAT episode number, which can loosely be conceived as the number of observed long term *not dispensed OAT* to *dispensed OAT* transitions. Other summaries, such as proportion of time dispensed OAT can seamlessly be used as an alternative stratification variable.

As pointed out by a referee, comparing our modelling with other methods that accommodate internal covariates, such as joint modelling and landmarking can provide insight towards our model's strengths and limitations. We defer the discussion between the modelling in this paper and joint modelling to a manuscript based on a manuscript based on Chapter 4 of [30]. In terms of landmarking however, there are two reasons why we have not adopted the landmarking framework: (i) researchers have shown joint modelling to out perform landmarking [24,28]; and (ii) Figure 1 illustrates an individual's OAT dispensation history is dynamic, and it is hence difficult to identify high quality landmark time(s).

The same referee also identified an extension to our proposed modelling by including a frailty term in (2), (3), and (4). The modelling presented in this paper explicitly assumes that the stratification variable $g(\mathcal{Z}(t))$ serves as an adequate summary of the covariate process up to time t. We have conducted some preliminary simulation studies which demonstrates our proposed modelling to perform well as long as $g(\mathcal{Z}(t))$ can summarize the majority of the process. In other words, our modelling is robust to this heterogeneity as long as the variance of the frailty term is small.

Although our methodology applies to an alternating binary process, our modelling is able to accommodate for more generally formulated treatment processes. For example, consider $Z(t) \in \{0, 1, \ldots, K\}$ with $K \geq 2$. In terms of our data application, the different levels of $Z(t)$ could be defined in terms of OAT type (methadone vs. buprenorphine-naloxone), dosage levels, and other relevant factors pertaining to OAT usage, such as receiving treatment under medical supervision. Conducting such an analysis could allow practitioners to identify which form of OAT is best suited for a particular individual,

and the potential hazards or benefits of different dosing and administration guidelines governing treatment.

Disclaimer

All inferences, opinions, and conclusions drawn in this manuscript are those of the authors, and do not reflect the opinions or policies of the British Columbia Ministry of Health and the Data Steward(s). The authors declare no potential conflict of interests.

Acknowledgments

The authors thank two anonymous reviewers for helpful comments and suggestions.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The research was supported in part by the Canada Graduate Scholarship [grant number CGSD3-504916] to T. J. Thomson, and grants to X.J. Hu from the Natural Sciences and Engineering Research Council of Canada [grant number NSERC-JRGPIN262823] and B. Nosyk from the National Institutes of Health [grant number NIH-R01DA050629].

Data availability statement

The data that support the findings of this study are not publicly available. The data may be made available upon submission of a reasonable Data Access Request to the British Columbia Ministry of Health and the data steward(s), subject to restrictions. More information may be found at: https://www2.gov.bc.ca/gov/content/health/conducting-health-research-evaluation/data-access-health-data-central. Statistical code is available from the corresponding author upon reasonable request.

References

[1] S. Barbieri, S. Mehat, B. Wu, C. Bharat, C. Poppe, L. Jorm, and R. Jackson, *Predicting cardiovascular risk from national administrative databases using a combined survival analysis and deep learning approach*, Int. J. Epidemiol. 51 (2022), pp. 931–944.

[2] J. Beyersmann and M. Schumacher, *Time-dependent covariates in the proportional subdistribution hazards model for competing risks*, Biostatistics 9 (2008), pp. 765–776.

[3] C. Bharat, L. Degenhardt, T. Dobbins, S. Larney, M. Farrella, and S. Barbieri, *Using administrative data to predict cessation risk and identify novel predictors among new entrants to opioid agonist treatment*, Drug. Alcohol. Depend. 228 (2021), pp. 109091.

[4] British Columbia Ministry of Health [creator], *Discharge Abstract Database (Hospital Separations)*, British Columbia Ministry of Health [publisher] 2018, available at health.gov.bc.ca/data/.

[5] British Columbia Ministry of Health [creator], *Medical Services Plan Payment Information File*, British Columbia Ministry of Health [publisher] 2018, available at health.gov.bc.ca/data/.

[6] British Columbia Ministry of Health [creator], *PharmaNet*, British Columbia Ministry of Health [publisher] 2018, available at health.gov.bc.ca/data/.

[7] British Columbia Vital Statistics Agency [creator], *Vital Statistics Deaths*, British Columbia Ministry of Health [publisher] 2018, available at health.gov.bc.ca/data/.
[8] R. Brookmeyer and N. Abdalla, *Multistate models and lifetime risk estimation: Application to Alzheimer's disease*, Stat. Med. 38 (2019), pp. 1558–1565.
[9] L. Clair, H. Anderson, C. Anderson, O. Ekuma, and H. Prior, *Cardiovascular disease and the risk of dementia: A survival analysis using administrative data from Manitoba*, Can. J. Public Health 113 (2022), pp. 455–464.
[10] R. Cook, J. Lawless, and B. Xie, *Marker-dependent observation and carry-forward of internal covariates in Cox regression*, Lifetime. Data. Anal. 28 (2022), pp. 560–584.
[11] G. Cortese and P. Andersen, *Competing risks and time-dependent covariates*, Biom. J. 51 (2009), pp. 138–158.
[12] D.R. Cox, *Regression models and life tables (with discussion)*, J. R. Stat. Soc. Ser. B 34 (1972), pp. 187–220.
[13] A. Hinds, L. Lix, M. Smith, H. Quan, and C. Sanmartin, *Quality of administrative health databases in Canada: A scoping review*, Can. J. Public Health 107 (2016), pp. e56–e61.
[14] X. Hu, M. Lorenzi, J. Spinelli, S. Ying, and M. McBride, *Analysis of recurrent events with non-negligible event duration, with application to assessing hospital utilization*, Lifetime. Data. Anal. 17 (2011), pp. 215–233.
[15] K. Iron and D. Manuel, *Quality assessment of administrative data (QuAAD): An opportunity for enhancing Ontario's health data*, ICES Investigative Report, Toronto, Institute for Clinical Evaluative Sciences, 2007.
[16] J.D. Kalbfleisch and R.L. Prentice, *The Statistical Analysis of Failure Time Data*, New York, Wiley, 2002.
[17] S. Larney, N. Jones, M. Hickman, S. Nielsen, R. Ali, and L. Degenhardt, *Does opioid agonist treatment reduce overdose mortality risk in people who are older or have physical comorbidities? Cohort study using linked administrative health data in New South Wales, Australia, 2002–17*, Addiction, 2023, pp. 1–13.
[18] D. Lin, L. Wei, I. Yang, and Z. Ying, *Semiparametric regression for the mean and rate functions of recurrent events*, J. R. Stat. Soc. Ser. B 62 (2000), pp. 711–730.
[19] Ministry of Public Safety and Solicitor General [creator], *British Columbia Corrections Dataset*, British Columbia Ministry of Health [publisher]; 2018, available at health.gov.bc.ca/data/.
[20] B. Nosyk, J. Min, L. Pearce, H. Zhou, F. Homayra, L. Wang, M. Piske, D. McCarty, G. Gardner, W. O'Briain, E. Wood, P. Daly, T. Walsh, and B. Henry, *Development and validation of health system performance measures for opioid use disorder in British Columbia, Canada*, Drug. Alcohol. Depend. 233 (2022), pp. 109375.
[21] L. Pearce, J. Min, M. Piske, H. Zhou, F. Homayra, A. Slaunwhite, M. Irvine, G. McGowan, and B. Nosyk, *Opioid agonist treatment and risk of mortality during an opioid overdose public health emergency: A population-based retrospective cohort study*, BMJ 386 (2020), pp. m772.
[22] R. Prentice, B. Williams, and A. Peterson, *On the regression analysis of multivariate failure time data*, Biometrika 68 (1981), pp. 373–379.
[23] D. Rizopoulos, *Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data*, Biometrics 67 (2011), pp. 819–829.
[24] D. Rizopoulos, G. Molenberghs, and E. Lesaffre, *Dynamic predictions with time-dependent covariates in survival analysis using joint modeling and landmarking*, Bio. J. 59 (2017), pp. 1261–1276.
[25] T. Santo Jr, M. Clark B. Hickman, J. Grebely, G. Campbell, L. Sordo, A. Chen, L. Tran, C. Bharat, P. Padmanathan, G. Cousins, J. Dubouy, K.E. R. Muga, B. Nosyk, J. Min, R. Pavarin, M. Farrell, and L. Degenhardt, *Association of opioid agonist treatment with all-cause mortality and specific causes of death among people with opioid dependence: A systematic review and meta-analysis*, JAMA. Psychiatry. 78 (2021), pp. 979–993.
[26] D. Schaubel, A. Wolfe, C. Sima, and R. Merion, *Estimating the effect of a time-dependent treatment by levels of an internal time-dependent covariate: Application to the contrast between liver wait-list and posttransplant mortality*, J. Am. Stat. Assoc. 104 (2009), pp. 49–59.

[27] L. Sordo, G. Barrio, M. Bravo, B. Indave, L. Degenhardt, L. Wiessing, M. Ferri, and R. Pastor-Barriuso, *Mortality risk during and after opioid substitution treatment: Systematic review and meta-analysis of cohort studies*, BMJ 357 (2017), pp. j1550.
[28] K. Suresh, J. Taylor, D. Spratt, S. Daignault, and A. Tsodikov, *Comparison of joint modeling and landmarking for dynamic prediction under an illness-death model*, Biom. J. 59 (2017), pp. 1277–1300.
[29] T.M. Therneau, *A Package for Survival Analysis in S*, 2015, available at https://CRAN.R-project.org/package=survival, version 2.38.
[30] T. Thomson, *Statistical methods for evaluating effects of health service utilization on survival times with linked administrative data, with application to opioid use disorder treatment management*, Ph.D. diss., Simon Fraser University, 2023.
[31] A. Tsiatis and M. Davidian, *Joint modeling of longitudinal and time-to-event data: An overview*, Stat. Sin. 14 (2004), pp. 809–834.
[32] H. Van Houwelingen, *Dynamic prediction by landmarking in event history analysis*, Scand. J. Stat. 34 (2007), pp. 70–85.
[33] Y. Zheng and P. Heagerty, *Partly conditional survival models for longitudinal data*, Biometrics 61 (2005), pp. 379–391.
[34] R. Zhou and J. Sun, *Estimation of the proportional mean residual life model with internal and longitudinal covariates*, Stat. Biosci. 14 (2022), pp. 550–563.