Evaluating Effects of Various Exposures on Mortality Risk of Opioid Use Disorders with Linked Administrative Databases

Trevor J. Thomson[1] · X. Joan Hu[1] · Bohdan Nosyk[2,3]

Abstract

Administrative health records provide a rich source of information pertaining to various exposures, many of which are time-varying in nature. When internal time-varying covariates are included in a Cox regression model, likelihood-based inference procedures are no longer applicable to infer model parameters (Kalbfleisch and Prentice in The Statistical analysis of failure time data, Wiley, New York, 2002). Motivated by the ongoing opioid epidemic, we summarize an individual's opioid agonist treatment (OAT) dispensation history and additional exposures with (i) a model-based summary, or (ii) its functional principal component scores. We show that the OAT dispensation proportion has a non-linear effect on the mortality hazard over time, and a significant interaction with time of birth. Particularly a clear protective effect against mortality for Millennials and Generation Z is revealed. Our approach is easy to implement by virtually any statistical software, and provides a risk assessment tool for utilizing available health records.

Keywords Cox regression · Functional principal components · Internal covariate · Stratified analysis

X. Joan Hu and Bohdan Nosyk have equally contributed to this work.

✉ Trevor J. Thomson
   trevor_thomson@sfu.ca

   X. Joan Hu
   joanh@stat.sfu.ca

   Bohdan Nosyk
   bohdan_nosyk@sfu.ca

1   Statistics and Actuarial Science, Simon Fraser University, Burnaby, BC, Canada

2   Centre for Health Evaluation & Outcome Sciences, St. Paul's Hospital, Vancouver, BC, Canada

3   Faculty of Health Sciences, Simon Fraser University, Burnaby, BC, Canada

🖄 Springer

1 Introduction

Administrative databases have become an increasingly popular data source to conduct population-based health research due to the availability of a vast collection of longitudinal clinical factors. The collected information allows researchers to investigate the association between time-varying exposures and some clinically meaningful event times, such as an individual's mortality time [1, 2, 7] Clinicians can use, for example, survival predictions to assess courses of treatments to meet their patient's specific needs.

The problem is that an individual with a health record on a particular date implies their survival (up to that date), which means that time-varying exposures are examples of internal time-varying covariates (cf. [13]). The challenge brought upon internal time-varying covariates is that the conventional relationship between the hazard and survivor functions no longer exists, which makes survival prediction problematic. However, given the vast information available in administrative databases, there is a need for survival prediction that utilizes these longitudinal factors [29, 31]. The conventional approach is to replace internal time-varying covariates with a summary of its observed history, and include them as covariates in a Cox regression model. For example, researchers have summarized an individual's observed ICD-9/10 codes (over time) to produce comorbidity scores (cfs. [16, 22, 23, 26]). Another approach is to replace the original covariate process with simpler processes, such as time-varying indicator variables (cf. [29]). Depending on the application however, using condensed histories of the exposure history can obscure the relationship between the exposure process and the event time of interest.

Motivated by the ongoing opioid epidemic, our aim is to predict survival probabilities given an individual's dispensation history of opioid agonist treatment (OAT), a prescribed treatment for opioid use disorders. Administrative health records between 01/01/1996 to 10/01/2018 in British Columbia, Canada in the form of drug dispensations [5], hospital and emergency department admissions [3], physician billing records [4], incarceration records [15], and deaths [6] were provided for individuals identified with an opioid use disorder with the objective to improve overall quality of care [18]. These health records not only provides us a real-world setting to generate scientific evidence on the clinical management of an opioid use disorder, but allow us to observe many time-varying exposures, a limitation with other prior studies [21].

Individuals were classified to be dispensed OAT or not dispensed OAT based on pharmaceutical records. In total there are 54,739 adults with at least one OAT dispensation record, in which the average OAT dispensation proportion across individuals is 58.37%. This highlights an individual's dispensation indicator to be a dynamic process, and a summary that utilizes the entire observed dispensation history is warranted. We use an individual's "overall OAT dispensation proportion" to summarize the entire dispensation process. In principle, this quantity is easy to compute, and can be included as a covariate in a Cox regression model. This variable however is generally unknown since complete dispensation

histories are unavailable due to data limitations. Such examples include the event time being subject to right-censoring, or data availability restricted to specific dates. To overcome this problem, we estimate it with the available data, and use it in place of the unknown quantity. However, using only the OAT dispensation proportion to summarize the entire covariate process may not adequately capture all relevant characteristics in terms of its effect on an individual's mortality hazard. Similar to [25], we apply functional principal component (FPC) analysis to the original exposure process in order to perform a dimensionality reduction, and summarize its history information with FPC scores (FPCSs). This approach allows the covariate process to self-identify relevant features from its history, without the need to explicitly specify a model. To the best of our knowledge, these summaries pertaining to OAT are not present within the literature, and would provide a new tool to assess the OAT effect on mortality. An added benefit of the proposed modeling is its ease to implement in statistical software, due to the development of an R package to obtain FPCSs with an alternating binary process [33].

The rest of the article is organized as follows. Sections 2 and 3 provides the Cox regression model and summaries of internal covariate processes, respectively. We apply the proposed modeling to the administrative database that motivated this research, and interpret our findings in Sect. 4. Finally, Sect. 5 provides some concluding remarks, and motivate the need for future investigation.

2 Notation and Modeling

Let T denote an individual's survival time (measured in *days*) since their first recorded OAT dispensation record. Suppose observations on T are subject to right-censoring with the censoring time, C. The available information on T is (T^*, Δ), where $T^* = T \wedge C$ is the minimum between T and C, and $\Delta = I(T \leq C)$. Let $Z(t) \in \{0, 1\}$ denote an individual's OAT dispensation indicator at time $t \geq 0$, which is obtained from daily pharmaceutical dispensation records, and $Z^H(t) = \{Z(u) : 0 \leq u \leq t\}$. We additionally let $X(t) = (X_1(t), \ldots, X_q(t))'$ denote external time-varying covariates and time-independent characteristics of an individual, and $X^H(t) = \{X(u) : 0 \leq u \leq t\}$. Here, the covariate $X_k(t)$ is time-independent if $X_k(0) \equiv X_k(t)$, for all $t > 0$, and $k = 1, \ldots, q$. Our statistical goal is to estimate the conditional hazard function of T (at time t), given the processes $Z^H(t)$ and $X^H(t)$:

$$\lambda\big(t; Z^H(t), X^H(t)\big) = \lim_{dt \to 0^+} \frac{1}{dt} P(t \leq T < t + \Delta t | T \geq t, Z(t), X(t)). \quad (1)$$

If a secondary objective is to obtain survival predictions based on a model for (2), we cannot directly use $Z^H(t)$ in our modeling, as this would obstruct the conventional relationship between the hazard and survivor functions. This relationship can be preserved if we can summarize an individual's entire history $Z^H(\infty) \equiv Z^H(T)$. Let ν denote a summary of $Z^H(\infty)$; we discuss how to obtain such a summary in Sect. 3. The key assumption is for ν to adequately summarize $Z(\infty)$, in the sense that T and

$Z^H(\cdot)$ are conditionally independent given ν. We then model the relationship between $(Z^H(\cdot),\ X^H(\cdot))$ and T by specifying the following Cox regression model with time-varying effects:

$$\lambda\big(t; Z^H(t), X^H(t)\big) \ = \ \lambda_0(t)\,\exp\{\gamma(t)\nu + \theta' X(t)\}. \tag{2}$$

Here, $\lambda_0(t)$ is an arbitrary baseline hazard function, $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_q)'$, and $\gamma(t)$ is an unknown function of time. One may approximate $\gamma(\cdot)$ with a linear combination of natural cubic spline basis functions with percentile-based knots [9],

$$\gamma(t) \approx \sum_{k=0}^{s} \phi_k C_k(t).$$

Here, s determines the number of interior knots in the spline. We use standard model selection criteria such as Akaike information criterion (AIC) or Bayesian information criterion (BIC) values to select an appropriate values for s [20].

With ν known, and assuming that the study collects n independent and identically distributed realizations of $\{(T_i^*,\ \Delta_i,\ \nu_i,\ X^H_i\ (T_i^*)): i=1, \cdots, n\}$, we estimate parameters $\boldsymbol{\theta}$ and $\gamma(t)$ by maximizing its partial likelihood function, and estimate $\Lambda\ (t) = \int_0^t \lambda_0(s)ds$ with the well-known Breslow estimator [14]. This allows us to estimate the survivor function with

$$\hat{S}\big(t; Z^H_i(t), X^H_{\ i}(t)\big) \ = \ \exp\{-\int_0^t \exp\left\{\widehat{\gamma}(t)\nu_i \ + \ \boldsymbol{\theta} X^H_i(t)\right\} d\hat{\Lambda}_0(u)$$

In practice however, we do not necessarily observe $Z^H(\infty)$, due to for instance, observing $Z(\cdot)$ only at times that align with a data extraction window. This renders ν to be unknown, and thus we cannot directly fit the model in (2). Viewing records of an internal covariate as longitudinal measures of an exposure process, the standard approach is to specify a model for the longitudinal outcomes that involves ν, and conduct so-called "joint modelling" with (2) [28]. This approach entails specifying the distribution of ν, and evaluating the likelihood function entails integrating over the distribution of ν. Prior research has shown the resulting estimator under to be fairly robust against misspecifying ν to follow a normal distribution (e.g. [10, 24, 27]), while attempts to reduce the computational costs with joint modeling (e.g. [8]) cannot avoid the integrating over the distribution of ν. This result makes implementing joint models a challenge, especially whenever the dimension of ν is large. This motivates us to consider an alternative approach, where we use the available data, $Z^H(T^*)$, to obtain our "best guess" for ν, say $\hat{\nu}$. We outline two alternative procedures to obtain $\hat{\nu}$.

Remark Due to the dispensation records only capturing records within British Columbia during the data extraction window, the date of the first *recorded* OAT dispensation may not necessarily correspond to an individual's *actual* first OAT dispensation. In other words, the time $t=0$ may not be meaningful for some individuals. We can overcome this issue by alternatively considering *age* as the timescale in (2), which transforms t to $a(t)=t+a_0$, where a_0 is an individual's age (in days) on

their first recorded OAT dispensation date. We consider both timescales within our data analysis in Sect. 4.

3 Summarizing Internal Covariates

We present two dimension reduction strategies for an individual's exposure history. The first approach summarizes the exposure history with a model-based estimate, and the second approach is to use the resulting FPCSs from the observed exposure history.

3.1 Average Dispensation Proportion

Let $R_i(t) = \int_0^t Z_i(u)du/t$ denote the proportion of time individual i is dispensed OAT over $[0, t]$ for $i = 1, \cdots, n$. In practice, one may choose to use $\{R_i(t_{ij}): j = 1, \cdots, m_i\}$, where $0 \le t_{i,m_i} \le T_i^*$ in their analysis. Using the simplified notation $R_{ij} = R_i(t_{ij})$, we specify the following model for the OAT dispensation history of individual i:

$$h(R_{ij}) = \nu_i + \varepsilon_{ij}, \tag{3}$$

for $j = 1, \cdots, m_i$ and $i = 1, \cdots, n$, where $h(\cdot)$ is some pre-specified function, ν_i is a subject-specific (unknown) quantity that summarizes individual i's OAT dispensation history, and ε_{ij} is a mean-zero term that captures the deviation between ν_i and the transformed measurements $h(R_{ij})$. We view ν_i as an unknown parameter in (3), with $E\{h(R_{ij})\} = \nu_i$ for all j. This naturally leads us to adopt $\hat{\nu}_i = \sum_{j=1}^{m_i} h(R_{ij})/m_i$ as an unbiased estimator for ν_i, regardless of the correlation between the ε_{ij}'s.

Although the model in (3) explicitly assumes that an individual's OAT dispensation proportion is constant with respect to time, this assumption can be relaxed by, for instance, replacing ν_i with linear combinations of natural cubic spline basis functions [30]. Although simplistic, the current specification of (3) allows us to interpret $\hat{\nu}_i$ as the *(observed) average of transformed proportions individual i was dispensed OAT during their follow-up time*. Furthermore, obtaining $\hat{\nu}_i$ is computationally simple, and it avoids the heavy computational costs present in joint modeling.

In order for $\hat{\nu}_i$ to serve as an adequate summary of ν_i however, we require (3) to be correctly specified, and m_i to be sufficiently large for each i. As an alternative approach, we adopt FPCSs to summarize the (observed) history of the covariate process. This approach avoids any model specification and pools all of the available information together, so it can therefore serve as an alternative method to a model-based summary.

3.2 Functional Principal Component Scores

A drawback with specifying a model to summarize the history of a time-varying exposure, as in Section 3.1, is that we may omit important features of the covariate process. This can happen if the model (3) is misspecified. This motivates us to view

the summary of the process as infinite-dimensional, and let the data identify important features of the process by conducting a dimension reduction procedure. As an individual's dispensation status over time can be viewed as a realization of a functional binary process, we applied sparse logistic FPC analysis for binary data [33], and use the FPCSs that explains a sufficient amount of variance (e.g. 95%) within the process, as summaries. The disadvantage of using principal scores as summaries lies with its interpretation, as it can be difficult to interpret the important features that are being summarized. In terms of its use in practice, Zhong and Zhang (2022) developed the R package SLFPCA to obtain FPCSs for alternating binary processes.

4 Analysis of the OAT Dataset

With the provincial administrative database, we fit the model in (2) to summarize the internal covariates. By using the method described in Section 3.1, we specified $h(x) = x$, so that the effect of interest pertains to the *average OAT dispensation proportion*. Additional risk factors we included in our analysis was based on a preliminary analysis conducted by [18]. Time-independent covariates include *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 recorded OAT dispensation date: 1996–2000 vs. 2001–2006 vs. 2007–2012 vs. 2013–2018). Although 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.

To summarize time-varying exposure processes, we started with one-jump binary processes to indicate if an individual satisfied a condition by time t. Since observing a particular value from these indicators does not inform us of an individual's survival status (except at the one time point where the indicator changes its value), we can loosely regard these indicators as external time-varying covariates. The time-varying covariates we consider are *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). The reason why we merged comorbidities with the term "*or* " in their names was to address the serial correlation induced when we initially treated all of the processes separately. Finally, the two internal covariates we proceed to summarize are *OAT dispensation indicator*, and *incarceration status*. The effect pertaining to *OAT dispensation indicator* is the effect of primary interest.

Table 1 presents the estimates of $\gamma(t) \equiv \gamma$ and $\boldsymbol{\theta}$ under (2), where the time scale is time since first recorded OAT dispensation and we used the average proportion and FPCSs as summaries for the OAT and incarceration histories. To conduct dimension reduction with the OAT process, we settled with the first two FPCSs, as 96.0% of

Table 1 Parameter estimates under the Cox regression model, where the time scale is time since first recorded OAT dispensation

	Model I: AIC=134,223.6; BIC=134,347.0		Model II: AIC=134,238.9; BIC=134,362.3		Model III: AIC=134,267.7; BIC=134,397.9		Model IV: AIC=134,278.2; BIC=134,408.5	
	Estimate	S.E	Estimate	S.E	Estimate	S.E	Estimate	S.E
Average OAT Dispensation Proportion	0.2607	0.0380	0.2471	0.0379	–	–	–	–
OAT FPC Score 1	–	–	–	–	0.0267	0.0129	0.0272	0.0129
OAT FPC Score 2	–	–	–	–	– 0.0049	0.0092	– 0.0050	0.0092
Average Incarcerated Proportion	0.6912	0.1391	–	–	0.6134	0.1408	–	–
Incarceration FPC Score	–	–	0.0283	0.0072	–	–	0.0281	0.0072
Sex (vs. Female)	–	–	–	–	–	–	–	–
Male	0.1681	0.0262	0.1842	0.0260	0.1754	0.0262	0.1894	0.0260
Birth Generation (vs Greatest & Silent Generations)								
Baby Boomers	– 1.3201	0.0440	– 1.3143	0.0440	– 1.3248	0.0441	– 1.3196	0.0441
Generation X	– 2.0861	0.0476	– 2.0660	0.0473	– 2.1083	0.0475	– 2.0893	0.0472
Millennials & Generation Z	– 2.1810	0.0591	– 2.1523	0.0587	– 2.1985	0.0591	– 2.1721	0.0587
Heath Authority (vs Fraser Health)	–	–	–	–	–	–	–	–
Interior	0.2030	0.0387	0.2048	0.0387	0.1951	0.0387	0.1968	0.0387
Vancouver Coastal	0.1080	0.0302	0.1058	0.0302	0.1078	0.0302	0.1059	0.0302
Vancouver Island	0.0803	0.0361	0.0806	0.0361	0.0842	0.0361	0.0842	0.0361
Northern	0.0405	0.0678	0.0499	0.0677	0.0315	0.0677	0.0396	0.0677
Year Category (vs. 1996–2000)	–	–	–	–	–	–	–	–
2001–2006	0.1646	0.0331	0.1629	0.0331	0.1480	0.0330	0.1472	0.0330
2007–2012	0.2502	0.0391	0.2432	0.0391	0.2392	0.0390	0.2337	0.0390
2013–2018	0.5871	0.0492	0.5914	0.0491	0.5793	0.0492	0.5829	0.0491
Alcohol or Other Substance Use Disorders	0.3525	0.0423	0.3588	0.0423	0.3712	0.0423	0.3759	0.0422
Ill Mental Health or Chronic pain	– 0.1782	0.0431	– 0.1812	0.0431	– 0.1817	0.0431	– 0.1844	0.0431
Hepatitis C Virus or HIV/AIDS	1.1486	0.0273	1.1533	0.0273	1.1529	0.0273	1.1569	0.0273

Table 1 (continued)

	Model I: AIC = 134,223.6; BIC = 134,347.0		Model II: AIC = 134,238.9 BIC = 134,362.3		Model III: AIC = 134,267.7; BIC = 134,397.9		Model IV: AIC = 134,278.2; BIC = 134,408.5	
	Estimate	S.E	Estimate	S.E	Estimate	S.E	Estimate	S.E
ver Received a Sedative	**0.4847**	0.0332	**0.4866**	0.0332	**0.4849**	0.0332	**0.4866**	0.0332
Ever on PharmaCare Plans C or G	**− 0.2276**	0.0301	**− 0.2174**	0.0300	**− 0.1844**	0.0295	**− 0.1772**	0.0294

Parameter estimates and test statistics that are bold-faced are statistically significant with 5% as the type 1 error rate. The model estimates in each block correspond to summarizing the OAT dispensation and incarceration indicators with either the average proportion or FPCSs

the overall variability is captured based on the estimated eigenvalues. In terms of *incarceration status*, we only include the first FPCS score as it explains 94.1% of the overall variability. In general, the signs of the estimates align with our expectation from descriptive statistics, with the nonconforming effects being *Mental Illness or Chronic Pain*, and *Ever on PharmaCare Plans C or G*. A possible reason for this can be due to these effects being correlated with the other time-varying indicator variables in the model. Illustrating the observed mean of these exposures over time [11] in Fig. 1 supports this hypothesis, and warrants the need of a more sophisticated summary beyond a one-jump binary as a topic for future investigation. Figure 1 also reveals that (i) non-survivors are generally more likely to receive exposures relative to survivors, and (ii) the pattern of exposures varies over time; a motivation for specifying a time-varying parameter in (2).

With regards to the effects of interest, we see that the effect of both *average OAT dispensation proportion* and *average incarcerated proportion* (an analogous summary of *incarceration status* from Sect. 3.1) are positive and statistically significant. We attribute the positive effect for *average incarcerated proportion* to nearly all non-survivors surviving during their incarceration, whereas the *average OAT dispensation proportion* result was surprising; we will further investigate this result later. However, as approximately 73.5% of individuals were never incarcerated, one may question the effectiveness of the *average incarcerated proportion* adequately summarizing the incarceration process, which motivated us to consider FPCSs. We examined the distribution of the FPCS for *incarceration status* in Table 2, in which the FPCSs appear to be uncorrelated with the *average incarcerated proportion*. Note that the functional principal component scores for the *OAT dispensation status* appear to be correlated with the *average OAT dispensation proportion*, and the results are similar between Model II and Model IV in Table 1. This led us to proceed with the covariate specification of Model II.
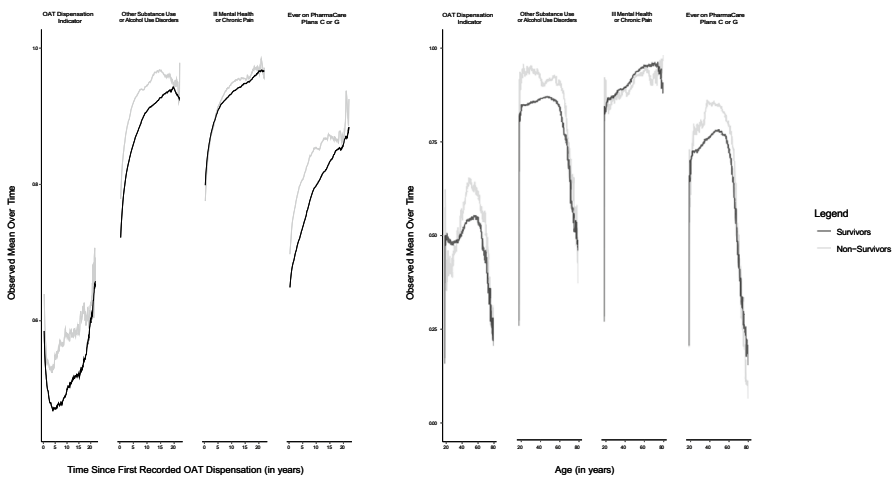

Fig. 1 Moving Average (30-days) of the observed mean over time (Hu and Lagakos 2007) for selected time-varying exposures, for all individuals stratified by their observed survival status

Table 2 Descriptive statistics for the FPCSs of OAT dispensation and incarceration records

OAT FPC Score 1

Average OAT dispensation rate > 0.6	Average OAT dispensation rate ≤ 0.6	Survivors	Non-survivors	Total
Minimum	− 21.6452	− 21.6452	− 16.8825	− 21.6452
1st Quartile	− 0.0163	− 0.0139	− 0.0330	− 0.0155
Median	0.0151	0.0196	0.0188	0.0195
3rd Quartile	0.0399	0.0524	0.0664	0.0542
Maximum	14.9780	17.7452	20.3875	20.3875
Mean	− 0.0690	− 0.0024	0.0164	0
SD	0.8018	0.9926	1.0488	1
N	26,178	47,731	7,008	54,739

Average OAT dispensation rate > 0.6 column:

	Average OAT dispensation rate > 0.6
Minimum	− 18.6848
1st Quartile	− 0.0131
Median	0.0392
3rd Quartile	0.0787
Maximum	20.3875
Mean	0.0633
SD	1.1485
N	28,561

OAT FPC score 2

Average OAT dispensation rate > 0.6	Average OAT dispensation rate ≤ 0.6	Survivors	Non-survivors	Total
Minimum	− 48.5866	− 48.5866	− 18.6510	− 48.5866
1st Quartile	− 0.0359	− 0.0481	− 0.0648	− 0.0497
Median	− 0.0126	− 0.0125	− 0.0134	− 0.0126
rd Quartile	0.0082	0.0143	0.0372	0.0161
Maximum	23.2177	23.2177	21.8200	23.2177
Mean	0.0635	0.0026	− 0.0180	0
SD	1.0113	0.9870	1.0843	1
N	26,178	47,731	7008	54,739

Average OAT dispensation rate > 0.6 column:

	Average OAT dispensation rate > 0.6
Minimum	− 19.9618
1st Quartile	− 0.0724
Median	− 0.0123
rd Quartile	0.0261
Maximum	19.0771
Mean	− 0.0582
SD	0.9859
N	28,561

Table 2 (continued)

OAT FPC Score 1

Incarceration FPC score

	Number of incarcerations = 0	Number of incarcerations > 0	Survivors	Non-survivors	Total
Minimum	− 0.0274	− 52.6459	− 52.6459	− 37.9442	− 52.6459
1st Quartile	− 0.0273	− 0.0279	− 0.0274	− 0.0274	− 0.0274
Median	− 0.0273	− 0.0273	− 0.0273	− 0.0273	− 0.0273
3rd quartile	− 0.0273	− 0.0259	− 0.0273	− 0.0273	− 0.0273
Maximum	− 0.0273	135.3005	135.3005	66.6001	135.3005
Mean	− 0.0273	0.0759	− 0.0006	0.0044	0
SD	<0.0001	1.9413	0.9989	1.0074	1
N	40,244	14,495	47,731	7008	54,739

To further investigate the effect estimate of *average OAT dispensation proportion*, we proceeded to fit the model in (2) to the observed data, where we used the *average OAT dispensation proportion* and *incarceration FPCS* as summaries. We specified $s \in \{0, 2, 3, 5\}$ for the number of interior knots in the spline, and used BIC to select the s; using AIC to select s produced similar results. Figure 2 illustrates the estimates and 95% confidence intervals for the *average OAT dispensation proportion* effect, where we included the estimate and 95% confidence interval under the Cox regression model, and the curve's average for reference. Here, the average of $\hat{\gamma}(\cdot)$ is $\int_0^t \gamma(t)dt/\tau$, where τ is the maximum domain of $\hat{\gamma}(\cdot)$. We see that the *average OAT dispensation proportion* effect is highly nonlinear, and is generally positive and statistically significant. The Cox regression estimate lying above the averaged curve value for time since first recorded OAT dispensation time scale is attributed to the (relatively) large death rate in the early stage of the study as opposed to the late stage, whereas the Cox estimate falls below the averaged curve with age as the time scale is due to a higher death rate in older individuals. As it is known that opioid tolerance varies with age [32, 35], this motivated us to explore the effect of *average OAT dispensation proportion* by first stratifying by birth generation. Table 3 shows the *average OAT dispensation proportion* across generations, where the difference of *average OAT dispensation proportion* between survivors and non-survivors clearly varies across birth generations. This justifies the need for birth generation specific effects, and suggests the results from Fig. 2 to be confounded by age group. Figure 3 shows the *average OAT dispensation proportion* effect for each generation for the two time scales we considered in Fig. 2. We see that the effects are generally linear (i.e. BIC generally selected $s = 0$) and the effects corroborate with the summary statistics in Table 3. In particular, the effect of *average OAT dispensation proportion* has a clear protective effect against mortality for *Millennials & Generation Z*.
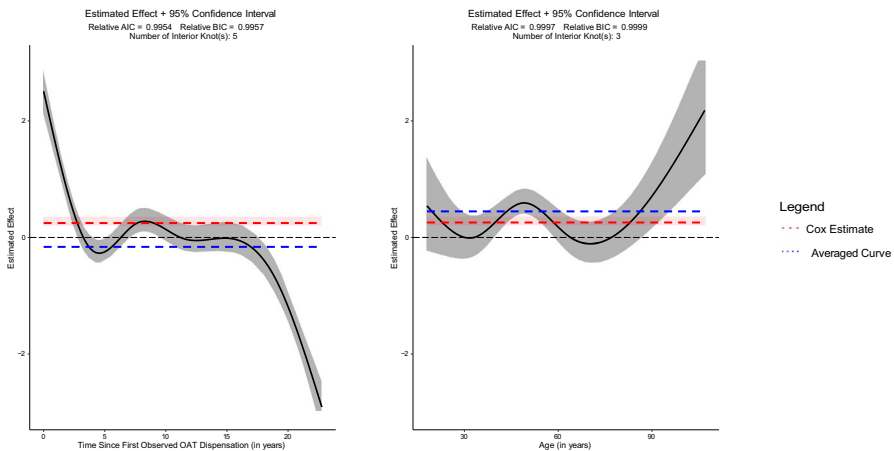

Fig. 2 Estimates and 95% confidence intervals for the *average OAT dispensation proportion*, where the time scales are time since first recorded OAT dispensation, and age. The average of the estimated curve and corresponding estimate under the Cox regression model are included as references

Table 3 Average OAT dispensation proportion summary statistics across birth generations

|	Survivors	Non-survivors	Total		Survivors	Non-survivors	Total
Greatest & silent generations: 1901–1945				Baby boomers: 1946–1964			
Minimum	0.0012	0.0014	0.0012	Minimum	0.0012	0.0022	0.0012
1st Quartile	0.1009	0.2539	0.1677	1st Quartile	0.2857	0.4067	0.3123
Median	0.4244	0.7972	0.6648	Median	0.6991	0.8005	0.7244
3rd Quartile	0.9272	0.9855	0.9756	3rd Quartile	0.9550	0.9704	0.9610
Maximum	1	1	1	Maximum	1	1	1
Mean	0.4900	0.6360	0.5768	Mean	0.6121	0.6731	0.6282
S.D	0.3836	0.3707	0.3826	S.D	0.3483	0.3271	0.3439
N	447	654	1,101	N	9,904	3,575	13,479
Generation X: 1965–1980				Millennials & Generation Z: 1981+			
Minimum	0.0013	0.0036	0.0013	Minimum	0.0014	0.0023	0.0014
1st Quartile	0.2750	0.2503	0.2727	1st Quartile	0.2631	0.1415	0.2574
Median	0.6146	0.5510	0.6076	Median	0.6001	0.4199	0.5929
3rd Quartile	0.9042	0.8322	0.8978	3rd Quartile	0.8998	0.7500	0.8951
Maximum	1	1	1	Maximum	1	1	1
Mean	0.5767	0.5375	0.5729	Mean	0.5693	0.4546	0.5647
S.D	0.3320	0.3174	0.3308	S.D	0.3335	0.3227	0.3338
N	18,861	1,998	20,859	N	18,519	781	19,300
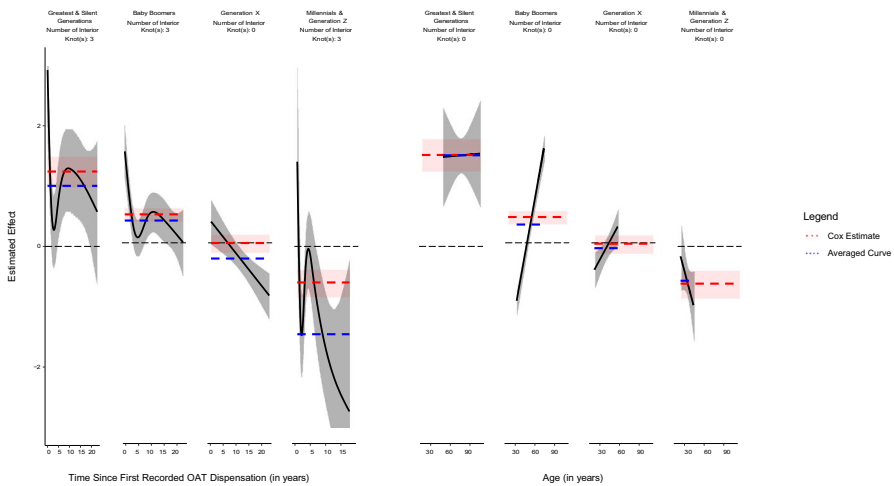
Fig. 3 Estimates and 95% confidence intervals for the average OAT dispensation proportion across each birth generation, where the time scale is specified as time since first recorded OAT dispensation, and age
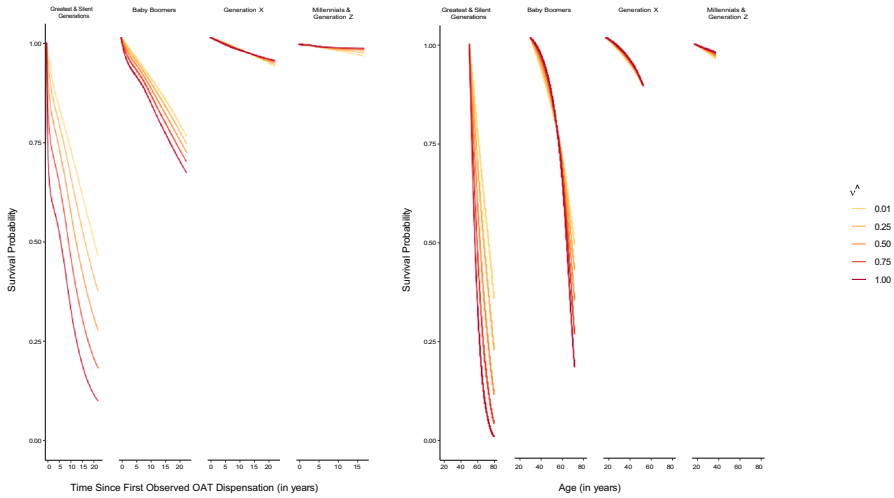
Fig. 4 Estimated survival probabilities for varying levels of $\hat{\nu}$, where the time scale is specified as time since first recorded OAT dispensation, and age. All other risk factors are held fixed at their reference level

By estimating the baseline hazard function with the well-known Breslow estimator, we illustrate estimated survival probabilities for various levels of $\hat{\nu}$ in Fig. 4. These survival probabilities were obtained by fixing all of the other risk factors at their reference level, and in particular, fixing the binary one-jump processes to zero. Similar to Fig. 3, we can see a clear protective effect of OAT for Millennials and Generation Z.
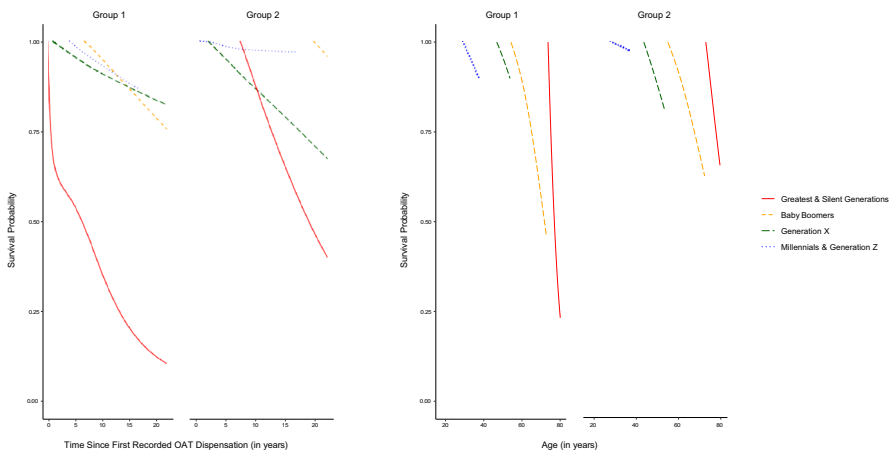
Fig. 5 Predicted survival probabilities for 8 randomly selected survivors (two from each of the four birth generation groups), where the time scale is specified as time since first recorded OAT dispensation, and age

Table 4 Information pertaining to the 8 randomly selected individuals with predicted survival probabilities illustrated in Fig. 5

	Group 1				Group 2			
Duration Between Study Dropout Date and First Recorded OAT Dispensation Date (in years)	0.0027	6.7151	0.8466	3.9753	7.3699	19.6137	2.0082	0.6521
Age at Study Dropout	73.6411	54.2027	46.8137	29.1753	73.3288	55.3178	43.9479	28.0877
Lifetime On OAT Proportion	1.0000	0.6142	0.7627	0.1143	0.1749	0.0198	0.0477	0.9952
Incarceration FPC Score	− 0.0274	− 0.0273	− 0.0274	− 0.0241	− 0.0273	− 0.0274	− 0.0273	− 0.0274
Sex	F	F	M	F	F	M	M	M
Birth Generation	G & S	BB	GX	M & GZ	G & S	BB	GX	M & GZ
Health Authority	F	VI	I	F	VI	VI	VC	F
Year Category	1996–2000	2007–2012	2013–2018	2013–2018	2007–2012	1996–2000	2013–2018	2013–2018
Alcohol or Substance Use Disorder Record[a]	N	Y	N	Y	Y	Y	Y	N
Ill Mental Health or Chronic Pain Record[a]	N	Y	N	Y	Y	Y	Y	Y
Hepatitis C Virus or HIV/AIDS Record[a]	N	N	N	N	N	N	N	N
Ever Received a Sedative[a]	N	N	N	N	Y	N	N	N
Ever on PharmaCare Plans C or G[a]	N	Y	N	Y	N	Y	Y	Y

F = female; M = male, G & S = greatest & silent generations, BB = baby boomers, GX = generation X, M & GZ = millennials & generation Z, FH = fraser health, I = interior, VC = vancouver coastal, VI = vancouver Island, N = no, Y = yes

[a]By the end of follow-up

We randomly selected 8 survivors (2 from each of the four birth generation groups) and predicted their survival probabilities based on their observed information, over time. We illustrate the survival probabilities in Figure 5, and provide information pertaining to these survivors in Table 4. As shown by the descriptive statistics in Table 3, individuals from earlier generations have a dramatically lower chance of survival relative to those from later birth generations. Although the effect of interest is the *average OAT dispensation proportion*, we can see from Table 4 that the covariates summarized with a one-jump binary process are correlated with each other, which was also reflected in Figure 1. Summarizing these processes while accounting for the correlation within these processes is certainly worthy to explore for future investigation, especially if the effect of these processes are of interest.

5 Final Remarks

Motivated by the demand of predicting an individual's mortality risk given their OAT dispensation history, this article directly uses time-invariant summaries of internal covariates as covariates in a Cox regression model. Since the entire history of internal covariates is generally unavailable, we model these summaries and estimate them from the available data. These summaries can be either specified as a function of the entire treatment process, such as an average from follow-up observations, or unspecified and allow the data to select important features of the covariate history. With health records provided by administrative data, we are able to incorporate several time-varying exposures within our modeling, and obtain an estimate that is closer to the true causal effect. Our results showcases the effect of OAT in our population to be non-linear and vary with respect to time of birth. In particular, we see a clear protective effect against mortality for Millennials and Generation Z. To the best of our knowledge, no other prior study has shown the effect of OAT to be non-linear or depend on an individual's birth generation.

To obtain personalized estimates, we considered stratifying individuals based on their birth generation. Rather than stratifying on a time-independent covariate, we can alternatively stratify individuals based on levels of a time-dependent covariate [12]. In particular, an upcoming publication directly uses $Z^H(\cdot)$ in a Cox regression model, where individuals stratification is time-varying and depends on levels of $Z^H(\cdot)$. Since that model directly uses $Z^H(\cdot)$, we cannot obtain survival predictions. A way to overcome this problem is to use a similar approach as this paper by summarizing internal covariates with ν, and base our stratification around ν.

The method we employ can essentially be summarized by conducting a Cox regression model, in which we replace the unknown covariate ν with an estimate $\hat{\nu}$. Although this approach is simple, we note that the estimated value can be expressed as $\hat{\nu} = \nu + \xi$, where ξ has mean zero. In other words, $\hat{\nu}$ can be seen as a "noisy measurement" of ν, and is ill-advised to directly replace ν [19]. Our analysis results align with descriptive statistics, and we hence believe that the induced bias is relatively small. An appropriate inference procedure to remove this bias, such as the conditional score approach [27] appears to be a promising approach to address this

issue, all while minimizing the computation intensity that likelihood-based procedures are known to suffer from.

Data Availability Access to data provided by the Data Steward(s) is subject to approval, but can be requested for research projects through the Data Steward(s) or their designated service providers. 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.

Declarations

Conflict of interest The authors declare no potential conflict of interests. X. Joan Hu is a co-editor of Statistics in Biosciences, but was not involved in the review process of this paper.

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).

References

1. Barbieri S, Mehat S, Wu B, Bharat C, Poppe C, Jorm L, Jackson R (2022) Predicting cardiovascular risk from national administrative databases using a combined survival analysis and deep learning approach. Int J Epidemiol 51(3):931–944
2. Bharat C, Degenhardt L, Dobbins T, Larney S, Farrella M, Bar-bieri S (2021) Using administrative data to predict cessation risk and identify novel predictors among new entrants to opioid agonist treatment. Drug Alcohol Depend 228:109091
3. British Columbia Ministry of Health: Discharge Abstract Database (Hospital Separations). British Columbia Ministry of Health (2018) British Columbia Ministry of Health. http://www.health.gov.bc.ca/data/
4. British Columbia Ministry of Health: Medical Services Plan Payment Information File. British Columbia Ministry of Health (2018) British Columbia Ministry of Health. http://www.health.gov.bc.ca/data/
5. British Columbia Ministry of Health: PharmaNet. British Columbia Ministry of Health, (2018). British Columbia Ministry of Health. http://www.health.gov.bc.ca/data/
6. British Columbia Vital Statistics Agency: Vital Statistics Deaths. British Columbia Ministry of Health (2018) British Columbia Ministry of Health. http://www.health.gov.bc.ca/data/
7. Clair L, Anderson H, Anderson C, Ekuma O, Prior HJ (2022) Car- diovascular disease and the risk of dementia: a survival analysis using administrative data from Manitoba. Can J Public Health 113:455–464
8. Ding J, Wang JL (2008) Modeling longitudinal data with nonparametric mul- tiplicative random effects jointly with survival data. Biometrics 64(2):546–556
9. Green PJ, Silverman BW (1994) Nonparametric regression and generalized linear models. Chapman and Hall, London
10. Hsieh F, Tseng YK, Wang JL (2006) Joint modelling of survival and lon- gitudinal data likelihood approach revisited. Biometrics 62:1037–1043

11. Hu XJ, Lagakos SW (2007) Nonparametric estimation of the mean function of a stochastic process with missing observations. Lifetime Data Anal 13:51–73

12. Hu XJ, Lorenzi M, Spinelli JJ, Ying SC, McBride ML (2011) Analysis of recurrent events with non-negligible event duration, with application to assessing hospital utilization. Lifetime Data Anal 17(2):215–233

13. Kalbfleisch JD, Prentice RL (2002) The statistical analysis of failure time data. Wiley, New York

14. Lin DY (2007) On the Breslow estimator. Lifetime Data Anal 13:471–480

15. Ministry of Public Safety and Solicitor General: British Columbia Corrections Dataset. British Columbia Ministry of Health (2018) British Columbia Ministry of Health. http://www.health.gov.bc.ca/data/

16. Needham DM, Scales DC, Laupacis A, Pronovost PJ (2005) A systematic review of the charlson comorbidity index using Canadian administra- tive databases: a perspective on risk adjustment in critical care research. J Crit Care 20(1):12–19

17. Nosyk B, Min JE, Pearce LA, Zhou H, Homayra F, Wang L, Piske M, McCarty D, Gardner G, Obriain W, Wood E, Daly P, Walsh T, Henry B (2022) Development and validation of health system per- formance measures for opioid use disorder in British Columbia, Canada. Drug Alcohol Depend 233(1):109375

18. Pearce LA, Min JE, Piske M, Zhou HC, Homayra F, Slaunwhite A, Irvine M, McGowan G, Nosyk B (2020) Opioid agonist treatment and risk of mortality during an opioid overdose public health emergency: a population-based retrospective cohort study. BMJ 386:772

19. Prentice R (1982) Covariate measurement errors and parameter estimates in a failure time regression model. Biometrika 69:331–342

20. Rice JA, Wu CO (2001) Nonparametric mixed-effects models for unequally sampled noisy curves. Biometrics 57:253–259

21. Santo T Jr, Clark B, Hickman M, Grebely J, Campbell G, Sordo L, Chen A, Tran LT, Bharat C, Padmanathan P, Cousins G, Dupouy J, Kelty E, Muga R, Nosyk B, Min J, Pavarin R, Farrell M, Degenhardt L (2021) 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. J Am Med Assoc Psychiatry 78(9):979–993

22. Schneeweiss S, Maclure M (2000) Use of comorbidity scores for control of con- founding in studies using administrative databases. Int J Epidemiol 29(5):891–898

23. Shin JH, Kunisawa S, Imanaka Y (2020) New outcome-specific comorbidity scores excelled in predicting in-hospital mortality and healthcare charges in administrative databases. J Clin Epidemiol 126:141–153

24. Song S, Davidian M, Tsiatis AA (2002) An estimator for the proportional hazards model with multiple longitudinal covariates measured with error. Biostatistics 3(4):511–528

25. Spreafico M, Ieva F (2021) Functional modeling of recurrent events on time-to-event processes. Biom J 63:948–967

26. Sun JW, Bourgeois FT, Haneuse S, Hernández-Díaz S, Landon JE, Bateman BT, Huybrechts KF (2021) Use of comorbidity scores for con- trol of confounding in studies using administrative databases. Am J Epidemiol 190(5):918–927

27. Tsiatis AA, Davidian M (2001) A semiparametric estimator for the propor- tional hazards model with longitudinal covariates measured with error. Biometrika 88(2):447–458

28. Tsiatis AA, Davidian M (2004) Joint modelling of longitudinal and time-to- event data: an overview. Stat Sin 14:809–834

29. Wong J, Taljaard M, Forster AJ, van Walraven C (2011) Does adding risk- trends to survival models improve in-hospital mortality predictions? A cohort study. BMC Health Serv Res 11(171):1–10

30. Wu L, Liu W, Hu XJ (2010) Joint inference on HIV viral dynamics and immune suppression in presence of measurement errors. Biometrics 66:327–335

31. Yamana H, Matsui H, Sasabuchi Y, Fushimi K, Yasunaga H (2015) Cat-egorized diagnoses and procedure records in an administrative database improved mortality prediction. J Clin Epidemiol 68(9):1028–1035

32. Zhao J, Xin X, Xie GX, Palmer PP, Huang YG (2012) Molecular and cel- lular mechanisms of the age-dependency of opioid analgesia and tolerance. Mol Pain 10(38):1–12

33. Zhong R, Liu S, Li H, Zhang J (2021) Sparse logistic functional principal component analysis for binary data. ArXiv

34. Zhong R, Zhang J (2022) SLFPCA: sparse logistic functional principal component analysis. R package version 2.0. https://CRAN.R-project.org/package=SLFPCA

35. Zubieta JK, Dannals RF, Frost JJ (1999) Gender and age influences on human brain mu-opioid receptor binding measured by PET. Am J Psychiatry 156:842–848