POST-PRINT: Abrahamowicz M, Beauchamp ME, Moura CS, Bernatsky S, Ferreira Guerra S, Danieli C. Adapting SIMEX to correct for bias due to interval-censored outcomes in survival analysis with time-varying exposure. *Biometrical Journal* 2022; 64(8):1467-1485. https://doi.org/10.1002/bimj.202100013

Adapting SIMEX to correct for bias due to interval-censored outcomes in survival analysis with time-varying exposure

Michal Abrahamowicz[*,1,2], Marie-Eve Beauchamp[2], Cristiano S. Moura[2], Sasha Bernatsky[1,2], Steve Ferreira Guerra[1], Coraline Danieli[2]

[1] Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, QC, Canada
[2] Centre for Outcomes Research and Evaluation, Research Institute of the McGill University Health Centre, Montreal, QC, Canada

Many clinical and epidemiological applications of survival analysis focus on interval-censored events that can be ascertained only at discrete times of clinic visits. This implies that the values of time-varying covariates are not correctly aligned with the true, unknown event times, inducing a bias in the estimated associations. To address this issue, we adapted the simulation-extrapolation (SIMEX) methodology, based on assessing how the estimates change with artificially increased time between clinic visits. We propose diagnostics to choose the extrapolating function. In simulations, the SIMEX-corrected estimates reduced considerably the bias to-the-null and generally yielded better bias/variance trade-off than conventional estimates. In a real-life pharmacoepidemiological application, the proposed method increased by 27% the excess hazard of the estimated association between a time-varying exposure, representing the 2-year cumulative duration of past use of a hypertensive medication, and the hazard of non-melanoma skin cancer (interval-censored events). These simulation-based and real-life results suggest that the proposed SIMEX-based correction may help improve the accuracy of estimated associations between time-varying exposures and the hazard of interval-censored events in large cohort studies where the events are recorded only at relatively sparse times of clinic visits/assessments. However, these advantages may be less certain for smaller studies and/or weak associations.

Key words: Cox model; Pharmacoepidemiology; Simulations; Time-varying covariates

Supporting Information for this article is available from the author or at https://doi.org/10.1002/bimj.202100013

1 Introduction

Classic methods of survival analysis rely on the assumption that the *exact* times of the occurrence of individual endpoint events of interest are known (Andersen et al., 2021). Yet, many applications of survival analysis in epidemiology and clinical research focus on endpoints for which the occurrence can be established only at the discrete times of clinical assessment, usually during a medical visit. Examples include cancer recurrence, the development of cognitive impairment, or the onset of AIDS. In such situations, it can only be ascertained that the event occurred within the interval between the last visit before and the first visit after its (unknown) true time of occurrence, implying interval-censored endpoints (Lindsey and Ryan, 1998; Zhang and Sun, 2010). Over the past four decades, many methods have been developed to extend survival analysis methods to such interval-censored outcomes (e.g., (Finkelstein, 1986; Turnbull, 1976; Kooperberg and Stone, 1992; Gentleman and Geyer, 1994)).

However, within the large body of literature on interval-censored data, relatively less attention was given to assessing the effects of *time-varying* covariates/exposures (TVCs/TVEs). For example, TVCs are *not* mentioned in some popular tutorials on interval-censored data (Lindsey and Ryan, 1998; Zhang and Sun, 2010; Sun et al., 2013). A recent PubMed search yielded 48 articles that mentioned both "interval censored" and "time-varying (or time-dependent)", including

[*]Corresponding author: Michal Abrahamowicz, Centre for Outcomes Research and Evaluation, Research Institute of the McGill University Health Centre, 5252 de Maisonneuve Blvd. West, Office 2B.38, Montreal, QC, H4A 3S5, Canada. Email: michal.abrahamowicz@mcgill.ca

statistical papers and real-life applications. Relevant statistical papers often proposed new methods to meet a specific analytical challenge encountered in analyses involving both TVCs and interval-censored events, e.g. stepped wedge clinical trials (Zhan et al., 2016), or TVCs representing the occurrence of a secondary interval-censored event (Ahn et al., 2018). Campbell et al. (2019) adapted joint modelling with shared random effects to model associations between TVCs and interval-censored outcomes, whereas Clark et al. (2014) adapted their imputation algorithm (Clark, Winchell and Betensky, 2013) to correct the inaccurate reported event times. Finally, Chen and Cook (2003) considered an interval-censored time-varying prognostic marker in the analysis of accurately timed recurrent events. Other authors extended earlier specialized methods to associations of TVCs with interval-censored outcomes, including the flexible HARE model (Kooperberg and Clarkson, 1997) and illness-death model (Chiang et al., 1989). Yet other authors proposed new tests for the association between a possibly interval-censored TVC and an interval-censored event time, but did *not* discuss multivariable modelling of such data (Schoenfeld et al., 2011; Oller and Gomez, 2020).

Several of the aforementioned papers do *not* report simulations (e.g., (Bacchetti and Quale, 2002; Sparling et al., 2006; Seaman and Bird, 2001; Clark et al., 2014)), making it difficult to assess how the proposed methods perform in plausible real-life situations. In addition, it was pointed that standard statistical packages do not allow the modelling of TVC associations with interval-censored events (Clark et al., 2014).

However, inaccurate timing of interval-censored events may have a strong impact on their estimated associations with TVCs, which require that, at each event time *t*, the corresponding hazard $\lambda(t)$ is correctly aligned with the updated TVC value $X_i(t)$ for each subject *i* in the risk set (Andersen and Liestol, 2003). Thus, if the true event times are distorted, the estimated regression coefficients will be based on incorrect TVC values and likely be biased. For example, we encountered this problem while assessing the impact of cumulative exposure to hydrochlorothiazide, a photo-sensitizing drug, on the hazard of non-melanoma skin cancer (NMSC) (see section 4). Cumulative exposure duration has to be modelled as a TVC and updated for each risk set (Andersen et al., 2021), whereas NMSC can be diagnosed only at visits to physicians with relevant specialties, i.e. is an interval-censored event. Thus, it is difficult to accurately establish for how long a given subject was truly exposed before his/her actual (unknown) event time. Yet, the literature provides little guidance regarding how this issue can be resolved in real-life applications. Indeed, our PubMed search failed to identify any pharmacoepidemiological study with time-varying drug exposure that attempted to account for interval-censored events. However, (i) effects of medications are often assessed using interval-censored endpoints, and (ii) time-varying exposures are essential to account for the considerable within-subject variation over time in drug use and/or its dose, and their potential cumulative effects (Patorno et al., 2015; Abrahamowicz, Beauchamp and Sylvestre, 2012; Pazzagli et al., 2018).

In this paper, we propose an easy-to-implement approach to account for interval-censored endpoints in Cox model-based analyses involving time-varying exposures, by adapting the simulation-extrapolation (SIMEX) methodology (Cook and Stefanski, 1994). Section 2 describes the proposed SIMEX-based procedure and related diagnostics. Section 3 presents simulations involving TVEs and an interval-censored endpoint. In section 4 we apply the proposed method to re-assess the association between HCTZ cumulative exposure and the NMSC hazard. The paper ends with a discussion of the implications of our results and limitations of our study.

2 Methods

2.1 Setting and objectives

We consider right censored time-to-event data, with $i = 1,\ldots,n$ independent subjects. For subject *i* the outcome is defined by $\{\tau_i, \delta_i\}$, where τ_i is the observed duration of follow-up and δ_i is the binary indicator of status at τ_i ($\delta_i = 1$ for event, $\delta_i = 0$ for censoring) (Andersen et al., 2021). We assume independent random censoring.

We focus on associations between a TVE $X(t)$ and the hazard of an interval-censored event. Specifically, $X_i(t)$ represents the updated value assigned to subject *i* at time *t.* Examples include the indicator of developing an infection at any time before *t*, or of the current drug use; an updated prognostic factor value; the current duration of a disease; and the cumulative dose of the drug received up to time *t*; each of which may be lagged to account for the latency in the exposure's impact on the current hazard (Danieli et al., 2019).

We assume that the association between $X_i(t)$ and the event time is analyzed using the multivariable Cox proportional hazards (PH) model (1), which adjusts for *p a priori* selected risk factors $Z_{i1}(t), \dots, Z_{ip}(t)$, which may include time-varying and time-invariant ($Z_{ij}(t) = Z_{ij}$, for all t) variables:

$$\lambda\left(t\middle|X_i(t), Z_{i1}(t), \dots, Z_{ip}(t)\right) = \lambda_0(t)\exp\left\{\beta_x X_i(t) + \sum_{j=1}^{p}\beta_j Z_{ij}(t)\right\} \quad (1)$$

In model (1), β's, the adjusted log hazard ratios (HRs), are estimated by maximizing the partial likelihood (PL) of the data, conditional on observed event times (Cox, 1972). PL is the product of the following components, each calculated for the risk set corresponding to one of the *m* observed events (for simplicity we assume no ties):

$$PL = \prod_{s=1}^{m}\left(\frac{\exp\{\beta_x X_s(\tau_s) + \sum_{j=1}^{p}\beta_j Z_{sj}(\tau_s)\}}{\sum_{i:\tau_i\geq\tau_s}\exp\left\{\beta_x X_i(\tau_s) + \sum_{j=1}^{p}\beta_j Z_{ij}(\tau_s)\right\}}\right) \quad (2)$$

where τ_s is the time of the s^{th} event, ordered in time, and all subjects still at risk at time τ_s (i.e. with $\tau_i \geq \tau_s$) are considered in the denominator.

Calculations in equation (2) rely on an implicit assumption that τ_s corresponds to the true time when the s^{th} event occurred. In contrast, in this manuscript we consider situations when the events are interval-censored, i.e. the exact event times are unknown. Specifically, we assume that assessing whether, for a given subject, the event has already occurred is possible only at discrete times, corresponding to his/her clinic visits or dates of specific diagnostic procedures (Lindsey and Ryan, 1998; Zhang and Sun, 2010). The timing and frequency of such "relevant visits" may vary between-subject and within-subject over time. Then, if the event for subject *s* is diagnosed/reported at visit time $t_{s,e}$, we can only deduce that the event occurred at some time between his/her most recent relevant visit $t_{s,e-1}$ and $t_{s,e}$. In real-life analyses of such events, the event times are imputed at τ_s^* typically corresponding to either the end ($\tau_s^* = t_{s,e}$) or the mid-point $\tau_s^* = \left(t_{s,e-1} + t_{s,e}\right)/2$ of the interval between the two relevant visits. Accordingly, the partial likelihood being maximized is redefined, relative to the classic definition in equation (2), as:

$$PL^* = \prod_{s=1}^{m}\left(\frac{\exp\{\beta_x^* X_s(\tau_s^*) + \sum_{j=1}^{p}\beta_{z,j}^* Z_{sj}(\tau_s^*)\}}{\sum_{i:\tau_i^*\geq\tau_s^*}\exp\left\{\beta_x^* X_i(\tau_s^*) + \sum_{j=1}^{p}\beta_{z,j}^* Z_{ij}(\tau_s^*)\right\}}\right) \quad (3)$$

where τ_s^* is the *imputed* time of the s^{th} event, $X_i(\tau_s^*)$ and $Z_{i1}(\tau_s^*), \dots, Z_{ip}(\tau_s^*)$ are the corresponding TVE/TVC values for subject *i* at τ_s^*, and β^*'s are the resulting log HRs.

Equation (3) uses different $X(t)$ and $Z(t)$ values than the classic formula (2) and these distortions affect the calculations for both "cases" (numerators of event-specific ratios) and "controls" (denominators) who remain at risk at τ_s^*. Moreover, in each risk set, the subsets of controls who contribute to the denominator for the s^{th} event in (3) vs. (2) may differ due to inaccurate imputation of event times in (3). For example, if the imputed time τ_s^* is later than the true event time τ_s, then the corresponding risk set used in (3) will incorrectly exclude subjects who are censored at any time c in between $\tau_s < c \leq \tau_s^*$. *Vice versa*, if the imputed time underestimates the true event time, the corresponding risk set in (3) may incorrectly include some subjects censored between times τ_s^* and τ_s. For these reasons, the coefficients $\hat{\beta}^*$'s that maximize *PL** in (3) will likely diverge from $\hat{\beta}$'s that would be estimated using *PL* in (2), based on true event times.

Our goal is to propose a pragmatic approach to reduce the impact of the aforementioned distortions of PL calculations on the accuracy of hazard ratios estimated for time-varying exposures.

2.2 Proposed SIMEX-like procedure

We propose to adapt the popular simulation-extrapolation (SIMEX) methodology to correct for errors due to interval-censored events in time-to-event regression analyses involving time-varying exposures (TVEs). SIMEX was

originally proposed for regression analyses involving a continuous independent variable X measured with additive random, normally distributed, measurement errors (MEs) $\epsilon_i \sim N(0, \sigma_\epsilon^2)$ (Cook and Stefanski, 1994). The classic SIMEX requires specifying the expected standard deviation σ_ϵ of MEs based on the literature, measurement calibration studies and/or validation analyses (Cook and Stefanski, 1994). Then, the main idea of SIMEX is to combine two steps: (i) adding further ME to the original mismeasured X_i, by simulation, to estimate how $\hat{\beta}_X$ varies with increasing ME, and (ii) extrapolate this relationship to the case of no ME ($\sigma_\epsilon^2 = 0$). Specifically, at consecutive iterations (j=1,...,k) of step (i), artificial data $X_i(\lambda_j)$ with additional measurement errors generated independently from $N(0, \sigma_\epsilon^2)$ are created as $X_i(\lambda_j) = X_i + \sqrt{\lambda_j} \cdot e_{ij}$, where $\lambda_j > 0$ is increased gradually across the iterations. This implies that the ME variance of modified covariate values $X_i(\lambda_j)$, at iteration j, is increased to $(\lambda_j + 1)\sigma_\epsilon^2$. At each iteration j, the regression model is re-estimated using the modified data $X_i(\lambda_j)$, and the corresponding coefficient $\hat{\beta}_{X,\lambda_j}$ is retained. At the extrapolation step (ii), the estimated values of $\hat{\beta}_{X,\lambda_j}$ are regressed on λ_j, using a simple user-defined parametric function $f(\lambda_j)$, e.g. linear, logarithmic, or quadratic (Cook and Stefanski, 1994). Finally, the estimated function $f(\lambda_j)$ is extrapolated to obtain $\hat{\beta}_{X,\lambda_j}$ corresponding to $\lambda_j = -1$, implying $(\lambda_j + 1)\sigma_\epsilon^2 = (-1 + 1)\sigma_\epsilon^2$, i.e. with no MEs, to obtain the SIMEX-corrected estimate of β_X. Notice that SIMEX does *not* attempt to improve the accuracy of the error-prone covariate measurements but, instead, to reduce the impact of MEs on the estimated regression coefficient (Cook and Stefanski, 1994).

In the specific setting described in section 2.1, the problem of inaccurate imputation of event times can be cast in general terms of MEs in the TVE $X(t)$. Indeed, the impact of using inaccurate information in PL calculations in equation (3) depends on how much individual $X_i(\tau_s^*)$ values at imputed event times diverge from the corresponding values $X_i(\tau_s)$ at the true, unknown, event times. However, it is difficult to quantify the expected standard deviation σ_ϵ of MEs $\{X_i(\tau_s^*) - X_i(\tau_s)\}$ across all subjects and risk sets. The ME for a given subject i at imputed event time τ_s^* will depend on both (a) the difference from the true event time $(\tau_s^* - \tau_s)$, and (b) the magnitude and pattern of within-subject changes over time in TVE values, which is difficult to quantify as it depends on the TVE metric used in the analyses (Abrahamowicz et al., 2012). For example, in pharmacoepidemiological applications, current exposure to drugs to treat chronic conditions will vary less than for drugs used only occasionally, or prescribed to treat short-term health problems. Furthermore, short-term exposure metrics, such a current daily exposure, will exhibit more within-subject variation, relatively, than cumulative long-term measures (e.g., total duration of drug use in the past 6 months).

To avoid the above difficulties, we propose an *ad hoc* modification of the SIMEX algorithm that circumvents the need to specify the standard deviation σ_ϵ of MEs in $X(t)$. Instead, we rely on a more easily quantifiable function of observed data that is expected to increase monotonically with increasing σ_ϵ. Specifically, our approach is based on the combination of two relationships. (a) On average, absolute ME $|X_i(\tau_s^*) - X_i(\tau_s)|$ should increase monotonically with increasing absolute distance between the two corresponding times $\Delta\tau = |\tau_s^* - \tau_s|$. (b) Furthermore, even if, in our context, the true event times τ_s are *not* known, mean $\Delta\tau$ will increase systematically with increasing mean difference between the visit times before ($t_{s,e-1}$) and after ($t_{s,e}$) the true event time, across all "cases" s = 1,…,m who had events diagnosed at their respective visit times $t_{s,e}$. Together, (a) and (b) imply that σ_ϵ is a monotonically increasing function of the mean distance between relevant visit times for cases:

$$\overline{\Delta t} = \sum_{s=1}^{m} (t_{s,e} - t_{s,e-1}) / m$$

Thus, we propose the following 4-step SIMEX-like procedure:

1. Through j = 1,…,k iterations gradually increase the mean time between visits $\overline{\Delta t}$ for "cases". Depending on the application, different approaches may be considered to simulate less frequent visit times. Section A.1 of the Supporting Information describes two alternative approaches that involve, respectively, (i) moving the time of previous relevant visits $t_{s,e-1}$ back by a pre-specified amount of time, or (ii) deleting some observed visits before $t_{s,e}$. Then, the imputed event time, for each case, is artificially moved to the mid-point of the interval between the *modified* time $t'_{s,e-1}$ of the last visit before the event and the observed time $t_{s,e}$ when the event was originally reported. In each iteration, a similar manipulation is applied to the time of the last visit for "controls" (who had no event during their follow-up), who are then right-censored at the corresponding modified time, because the event could not be observed after their last visit.

2. In each iteration j (j = 1,…,k), refit the Cox PH model of interest using the modified times of events for cases and censoring for controls (based on step 1). Retain the corresponding estimated log HR for exposure ($\hat{\beta}_{X,j}$).

3. Regress the k+1 point estimates, including the estimate from the original analysis of the unmodified data ($\hat{\beta}_{X,0}$ for j=0) and k estimates $\hat{\beta}_{X,j}$ from step 2, on the corresponding mean $\overline{\Delta t}_j$ of the difference between the (modified) times of the two relevant visits for cases. Here, we fit a simple linear regression model $\hat{\beta}_{X,j} = \alpha_0 + \alpha_1 f(\overline{\Delta t}_j) + \varepsilon$, where $f(\cdot)$ is a pre-specified parametric monotone function. (Diagnostics to help choose the functional form of $f(\cdot)$ are discussed in section 2.3.)

4. Use the regression model estimated in step 3 to extrapolate the expected value of log HR $\hat{\beta}_X$ to a hypothetical ideal scenario where the mean time between relevant visits is 1 time unit, i.e. $\overline{\Delta t} = 1$, implying no errors in the timing of events. The resulting extrapolated value is considered the SIMEX-corrected log HR estimate for the exposure.

To get the bootstrap-based 95% confidence interval (CI) for the SIMEX-corrected log HR, we repeat steps 1-4 of the SIMEX procedure in each of B bootstrap resamples, and use the percentile method, based on the normal approximation for the resulting distribution of SIMEX-corrected estimates.

The proposed SIMEX-like procedure shares the underlying rationale with and involves similar data manipulations as the DELEX (deletion and extrapolation) algorithm (Andersen and Liestol, 2003). However, DELEX focuses on correcting for sparse measurements of time-varying covariates, rather than for inaccurate times of interval-censored events.

Similar to other authors who extended the SIMEX methodology to different complex settings (e.g., (Kyle, Moodie and Abrahamowicz, 2016; Kuchenhoff, Mwalili and Lesaffre, 2006; Wang, Beauchamp and Abrahamowicz, 2020; Oh et al., 2018)), in section 3 we rely on simulations to validate the proposed method in our specific context of interval-censored events.

R programs used to run the simulations are available as Supporting Information. R code allowing end-users to implement our proposed SIMEX-like method for a given dataset and an illustrative example based on a simulated dataset are available at https://github.com/mebeauchamp/SIMEX-IntervalCensoredEvents. The code for producing the diagnostic plot to help choosing the extrapolating function and for estimating the bootstrap CI is also included.

2.3 Choice of the extrapolating function

The final SIMEX-corrected estimate depends on the choice of the functional form of the extrapolating function $f(\cdot)$ in steps 3 and 4. This issue, common to all SIMEX-based methods (Cook and Stefanski, 1994), becomes challenging for non-linear models (e.g., Cox model) where the optimal choice cannot be analytically derived, so *exploratory* analyses may offer insights (Stefanski and Cook, 1995). Below, we provide practical suggestions to help with this choice in real-life applications.

The shape of the function $f(\cdot)$ should describe how estimated log HR $\hat{\beta}_{X,j}$ for the TVE $X(t)$ changes, across the SIMEX k iterations (j = 1,…,k), with increasing mean difference ($\overline{\Delta t}_j$) between the (modified) times of the two relevant visits for cases. These changes in $\hat{\beta}_{X,j}$ will reflect the impact of increasing MEs in $X(t)$, i.e. discrepancies between exposure values at the imputed vs. the corresponding true event time. This impact, in turn, depends on both the individual longitudinal patterns of exposure (e.g., drug use) and properties of the chosen exposure metric. For example, if most users of the drug remained exposed for several months, without interruptions, then TVE values would be little affected by even large discrepancies between imputed vs. true event times. Yet, in most real-life applications, individual exposure patterns vary considerably between subjects and changes in $\hat{\beta}_{X,j}$ depend mostly on the average amount of resulting errors in the TVE. Still, more "acute" exposure metrics, such as current exposure or total dose from the past week, are relatively more affected by short-term changes in individual exposure status/dose than long-term metrics, such as cumulative dose since the beginning of follow-up. Thus, the choice of the extrapolating function should be approached as an empirical modelling problem (Lindsey and Ryan, 1998).

In our specific context, the full history of individual exposure values $X_i(t)$ is available in prescription databases used in most pharmacoepidemiologic studies (MacMahon and Collins, 2001; Abrahamowicz and Tamblyn, 2005; Avorn, 2007). Thus, one can assess how the discrepancy between an individual's exposure at different times $X_i(t)$ vs. $X_i(u)$ varies with increasing time differences $t - u$. Specifically, based on measurement error theory (Carroll et al., 2006), we consider the correlation $r\{X_i(t), X_i(u)\}$ as an indicator of the ability of the TVE value $X_i(u)$, at time u, to serve as a proxy for $X_i(t)$.

In addition, because the problem is directly related to inaccuracies in establishing event times, we focus on cases. For each case *s*, we first identify his/her imputed event time τ_s^* and the corresponding TVE value $X_s(\tau_s^*)$. Then, in p = *1,...,P* iterations, we gradually move the presumed *event* time back, by $p \cdot \delta t$ (where e.g. $\delta t = 1$ week or 1 month), and establish the corresponding exposure values: $X_s(\tau_s^* - \delta t)$, $X_s(\tau_s^* - 2\delta t)$, etc. For each *p,* we calculate the Pearson correlation coefficient between pairs of exposure values for the same case $r(p) = r\{X_s(\tau_s^* - p \cdot \delta t), X_s(\tau_s^*)\}$. Finally, we plot the correlations $r\{X_s(\tau_s^* - p \cdot \delta t), X_s(\tau_s^*)\}$ against the corresponding time discrepancies $p \cdot \delta t$. The resulting curve illustrates how the ability of TVE $X_s(\tau_s^*)$ to mimic its "target" value $X_s(\tau_s)$ varies with increased discrepancy between the corresponding times τ_s^* vs. τ_s, and, thus, provides an empirical basis to (i) select the analytical form of the SIMEX extrapolation function $f(\cdot)$, and (ii) help determine the number *k* of iterations for SIMEX step 1.

Notice that these diagnostics rely exclusively on *observed* data, and thus, can be implemented in real-life settings, as illustrated in our simulations in section 3. In all main simulations, summarized in Table 1, and for both TVE metrics, the correlation $r\{X_s(\tau_s^* - p \cdot \delta t), X_s(\tau_s^*)\}$ decreased, in an approximately *linear* way, with time discrepancy $p \cdot \delta t$ increasing from 0 to some relatively clear threshold $p' \cdot \delta t$ (see Figure 2 for 10 samples simulated for scenario 1 and Figure 3 for the skin cancer application (section 4)). Furthermore, for all main simulations, using a linear extrapolation function at SIMEX steps 3 and 4 resulted in lower bias and root mean squared error (RMSE) of estimates than using quadratic or logarithmic extrapolation functions (Table B.2 in the Supporting Information provides results for one scenario). In contrast, in additional simulations based on "change-of-status exposure" generation, which resulted in higher short-term within-subject variation in the binary indicator of current drug use, the correlation changed in a *non-linear* fashion in the low range of $p \cdot \delta t$ values (Figure B.1 of the Supporting Information, panel a)). This suggested a *logarithmic* function which, indeed, in these additional simulations, considerably reduced the bias in the estimated effect of *current use* (but *not* for cumulative duration of use in the past 12 weeks) relative to the linear extrapolation function (Table B.4 of the Supporting Information).

Based on these simulation results, we propose a 4-step approach to select the extrapolation function for a given real-life application:

a) Start with a visual assessment of the pattern of changes in the correlation $r(p) = r\{X_s(\tau_s^* - p \cdot \delta t), X_s(\tau_s^*)\}$ on the y axis, with increasing time difference $p \cdot \delta t$ between the two times when the TVE is assessed on the x axis. Identify on the x axis the value corresponding to the mean distance $\overline{\Delta t}$ between relevant visits in observed data, i.e. ¼ $\cdot \overline{\Delta t}$. (When imputing the event time at the mid-point of the interval between two visits, the expected value of the discrepancy from the true event time that occurred in the same interval equals about ¼ of the interval's length, assuming an approximately *uniform* distribution of true event times within the intervals.) In SIMEX step 1, the difference between the imputed and true event times will progressively be increased beyond ¼ $\cdot \overline{\Delta t}$ in successive iterations, and then, we must consider time differences $p \cdot \delta t$ beyond this point to inform about the extrapolating function $f(\cdot)$.

b) If *no* major departure from linearity – for some interval between 0 and $p' \cdot \delta t$ beyond ¼ $\cdot \overline{\Delta t}$ – is observed on the graph, choose the linear function as the default option.

c) If, however, the diagnostic graph shows an important non-linearity for values of $p \cdot \delta t$ from 0 to early after ¼ $\cdot \overline{\Delta t}$, choose a simple 1-degree-of-freedom (df) *non-linear* function, e.g. $\log(p \cdot \delta t)$ or $(p \cdot \delta t)^2$, whose shape approximates the curve.

d) Identify an approximate upper bound $p' \cdot \delta t$, beyond ¼ $\cdot \overline{\Delta t}$, over which the function chosen in step b) or c) mimics the corresponding $r(p)$. Then, choose the number *k* of iterations for SIMEX step 1 so that the upper limit $\overline{\Delta t}_k$ of the difference between the (modified) times of two relevant visits corresponds approximately to

4 times this threshold ($p' \cdot \delta t$). In sensitivity analyses, one may increase $p'$and, thus, the number of SIMEX iterations k, to assess the robustness of the SIMEX-corrected estimate.

However, as illustrated in real-life analyses in section 4, if the exposure metric is lagged to account for the latency (L) between the exposure and an event such as cancer occurrence (Danieli et al., 2019; Richardson and Ashmore, 2005), increasing the distance between relevant visits $\overline{\Delta t}_j$ in SIMEX iterations j=1,...,k entails also decreasing artificially the number of observed events. Indeed, all original events for which the modified imputed event times, at SIMEX iteration j, fall within the period (0, L] will not contribute to estimates for iterations j,…,k. Section C.4 and Figure C.1 of the Supporting Information provide further explanations. In such cases, to avoid (i) unstable estimates for later SIMEX iterations due to considerably reduced number of events, and (ii) serious incompatibility of exposure log HR estimates across iterations (which rely on different numbers of events), we suggest limiting the number of SIMEX iterations k so that the corresponding estimates in SIMEX step 2 rely on at least 80% of the originally observed events, and then ignore step d) above.

3 Simulations

3.1 Methods

Below, we describe the main features of our simulations using the ADEMP (aims, data-generating mechanisms, estimands, methods, performance measures) structure (Morris, White and Crowther, 2019).

Aims

Our simulations aim at assessing the performance of the SIMEX-based correction for interval-censored events, proposed in section 2, in hypothetical prospective or retrospective pharmacoepidemiological cohort studies of an adverse effect of a time-varying drug exposure. We assume that a complete drug exposure history can be reconstructed based on recorded timing and duration of prescriptions filled by individual patients (Abrahamowicz and Tamblyn, 2005; Patorno et al., 2015).

Data-generating mechanisms and assumptions

We generated cohorts of N=750, 1500, or 3000 new users of a drug (Ray, 2005), followed from time 0 until either an adverse event of interest or right censoring at the administrative end of the study, at 240 weeks. Without any loss of generalizability, no censoring on competing events or losses to follow-up were considered.

For each subject, we randomly generated age (uniform U(40, 80)) and, independently, sex with P(woman)=0.7. To generate the time-varying exposure (TVE), we first simulated N individual vectors of 240 binary values of current drug use in week t: $U_i(t)$, i = 1,…,N, t = 1,…,240. Different subjects had different propensity to use the drug, and its use was intermittent, with alternating periods of use and non-use, whose durations varied both within-subject over time and across subjects (Sylvestre and Abrahamowicz, 2009). Section B.1 of the Supporting Information provides details of two different approaches used to generate time-varying exposures. The "duration-based exposure" approach was used in the main simulations while the "change-of-status exposure" was employed in additional simulations. For each subject i, based on his/her generated $U_i(t)$ vector, we then calculated the corresponding time-varying values of the cumulative duration of use in the past 12 weeks $Cum12_i(t) = \sum_{r=t-12}^{t} U_i(r)$.

The events were generated, separately for $U(t)$ and $Cum12(t)$, from the exposure-specific PH model, with the hazard depending on the respective TVE and time-invariant values of age and sex:

$$\lambda_i(t|age_i, sex_i, U_i(t)) = \lambda_0 \exp\{0.05\, age_i + 0.8\, sex_i + \beta_U U_i(t)\}$$

$$\lambda_i(t|age_i, sex_i, Cum12_i(t)) = \lambda_0 \exp\{0.05\, age_i + 0.8\, sex_i + \beta_{Cum12} Cum12_i(t)\}$$

To generate either (i) stronger or (ii) weaker exposure effects, we assumed, respectively, (i) log HRs β_U = log(3) = 1.01 and β_{Cum12} = log(2)/6 = 0.12 for one week increase in *Cum12(t)*; or (ii) β_U = log(2) = 0.69 and β_{Cum12} = log(3)/12=0.09.

The combination of the three sample sizes with stronger/weaker effects resulted in six different main simulated scenarios for the duration-based exposure simulations (scenarios 1-6). Similar scenarios were replicated in additional change-of-status exposure simulations, reported in section B.6 of the Supporting Information (scenarios 9-14). In two further simulated scenarios (scenarios 7-8), with *N*=3000, duration-based exposure generation and strong effects, we assumed that (a) the hazard depends also on an additional time-varying binary risk factor generated independently of exposure, using duration-based approach, and (b) exposure has no effect (true HR=1.0).

For each scenario, the constant baseline hazard $\lambda_0$was defined so that about one third of the cohort had uncensored events in the 240 weeks of follow-up, resulting in approximately 67% censoring rate. We used the permutational algorithm, validated for simulating event times conditional on time-varying exposures (Sylvestre and Abrahamowicz, 2008; Sylvestre et al., 2015) to generate individual event times (see section B.2 of the Supporting Information for details).

Finally, assuming the adverse event is an *interval-censored* event, we generated individual vectors of consecutive irregular clinic visit times, when its occurrence could have been ascertained (see section B.3 of the Supporting Information).

For each simulated scenario, we generated 1000 independent and identically distributed random samples.

Methods

Each simulated sample was analyzed with a series of correctly specified Cox PH models, which included the TVE used to generate the data ($U(t)$ or $Cum12(t)$), age and sex, and, if relevant, the additional time-varying risk factor. The models differed only with respect to how they handled the event and censoring times.

First, we estimated two *conventional* Cox models that made *no* attempt to correct for the inaccurate (imputed) timing of events. In the first, for each patient with an *observed* event, the imputed event time corresponded to the mid-point of the interval between the patient's last visit before and his/her first visit after the true (generated but assumed unobserved) event time. Subjects were censored at the time of their last visit before 240 weeks of follow-up, if they had no events until then. (Thus, a few events, generated between the subject's last visit and 240 weeks, could not be "observed" and were ignored in the analyses.) Figure 1 illustrates how this naïve imputation, at the interval's mid-point, induces errors in the corresponding TVE values, relative to their values at the true event time. The second conventional model imputed event times at the *end* of the relevant interval.

The same data were then analyzed using the SIMEX-like procedure proposed in section 2.2, with the extrapolating function $f(\cdot)$ chosen based on diagnostics of section 2.3. For details of SIMEX implementation see section B.4 of the Supporting Information.

Finally, as a benchmark, we fit the "oracle" Cox model which used the *true* times of all events.

Estimands

For each model considered, the estimand of primary interest was the adjusted log HR ($\hat{\beta}_U$ or $\hat{\beta}_{Cum12}$) for the respective TVE.

Performance measures

Estimated exposure log HRs from the 1000 samples simulated per scenario were compared with the corresponding true exposure effect, used to generate events, and summarized with respect to: relative bias, empirical standard deviation (SD), and root mean squared error (RMSE). In limited additional simulations, we assessed empirical coverage rates of the 95% CI for $\hat{\beta}_U$ or $\hat{\beta}_{Cum12}$ obtained with different models, i.e. the proportion of simulated samples where the 95% CI included the true exposure log HR.

3.2 Simulation results

This section summarizes results for the main simulations with "duration-based exposure" generation (see section B.1 of the Supporting Information for details). The mean number of uncensored events was about 956, 478 and 240 in scenarios with *N*=3000, 1500 and 750, respectively. Section B.5 of the Supporting Information reports details on the distance between visits and exposure metrics in generated data.

To choose the extrapolation function used at steps 3 and 4 of the SIMEX-like approach, we relied on diagnostics proposed in section 2.3. Figure 2 shows, for 10 random samples of scenario 1 (*N*=3000, stronger exposure effects, and duration-based exposure generation), how the correlation $r(p) = r\{X_s(\tau_s^* - p \cdot \delta t), X_s(\tau_s^*)\}$, across all m cases, varies with increasing difference in time δt, relative to the imputed event time τ_s^*. For both exposure metrics (current use $U(t)$ and cumulative duration of use in the past 12 weeks $Cum12(t)$)), $r(p)$ increases approximately linearly for δt increasing from 0 to about 10-15 weeks, and then gradually stabilizes for larger time differences δt > 20 weeks. Based on these diagnostics, we decided to rely on the *linear* extrapolating function for SIMEX, and limited the range of δt values used to estimate this function to about 13 weeks. The mean expected error in the timing of an event *s*, imputed at the mid-point of the interval between the two relevant visits, is close to ¼ of the length of this interval $[t_{s,e-1}, t_{s,e}]$. Thus, in SIMEX iterations we aimed at about 52 weeks (4*13 weeks) as the upper bound for the range for the artificially increased mean difference between relevant visits for cases $\overline{\Delta t_j}$ (see section 2.3 for details). This implied *k*=15 iterations, with $\overline{\Delta t_j}$ increasing from $\overline{\Delta t}$ = 25.6 weeks for the "*observed"* data (vertical dotted lines in Figure 2, at 25.6/4 = 6.4 weeks) to $\overline{\Delta t}_{15}$ = 51.0 weeks (1st dashed vertical line in each panel in Figure 2, at 51.0/4=12.8 weeks). However, we recognize that in real-life applications, the exact number of iterations used to estimate the extrapolating function may be subject to arbitrary decisions. Therefore, as sensitivity analyses, we increased the number of iterations to *k*=25, with $\overline{\Delta t_j}$ distributed between 25.6 and 65.7 weeks (2nd vertical dashed lines in Figure 2, at 65.7/4=16.4 weeks).

Table 1 and Table B.1 in the Supporting Information compare the log HR estimates for the effects of (i) current use $U(t)$ (upper half) and (ii) cumulative duration of drug use in the past 12 weeks (lower half), obtained with different modelling strategies, for the six main simulated scenarios. As expected, the "oracle" analyses, which relied on true event times, yielded the lowest RMSEs and uniformly unbiased estimates (Supporting Information Table B.1), which indirectly validates our approach for simulating individual event times based on the permutational algorithm (Sylvestre and Abrahamowicz, 2008). The conventional estimates, which did not correct for inaccurately imputed event times at the midpoint of the relevant intervals, yielded relative underestimation bias of about 50% for current use and 30% for cumulative duration (Table 1). Biases of conventional estimates increased further to about 67% and 52%, respectively, if events were imputed at the end of the interval (Supporting Information Table B.1). The bias is stronger for the binary indicator of current use, for which an error in the imputed value implies a complete reversal of the true exposure status.

SIMEX estimates, with linear extrapolation, reduced the relative bias to 16%-25% for current use and only to 6%-10% for cumulation duration (Table 1). However, as expected, the SDs of SIMEX estimates were about twice higher than for conventional estimates (Table 1). Even so, in all simulations summarized in Table 1, the bias/variance trade-off, in terms of lower RMSE, was systematically better for SIMEX-corrected estimates. The benefits of SIMEX were most evident for larger *N* and/or stronger exposure effects (Table 1), where RMSE was more affected by the bias than by variance. Still, even with only *N*=750 and 239 events, for weaker exposure effects, SIMEX reduced RMSEs moderately, by 11%, for current use and slightly, by 5%, for cumulative duration (scenario 6 in Table 1). All aforementioned SIMEX results were similar for *k*=15 vs. *k*=25 iterations at step 1 (Table 1), confirming the robustness of our estimates.

Table 2 and Table B.3 of the Supporting Information summarize results for two variations of scenario 1. Scenario 7 shows that including an additional time-varying covariate does *not* change materially the results for exposure, confirming the applicability of the proposed SIMEX-based approach in multivariable analyses. Scenario 8 indicates that, as expected, when the TVE is *not* associated with the hazard (true β=0), both conventional and SIMEX estimates are unbiased, but the latter inflates the variance.

Similar results for additional simulations, with TVEs generated using the "change-of-status" approach are reported in Supplementary Tables B.4 and B.5, with brief comments in section B.6, of the Supporting Information.

Finally, we used 500 random samples simulated for scenario 3 to assess the coverage rates of 95% bootstrap CIs, based on 300 bootstrap resamples. Due to important biases (Table 1), conventional estimates yielded very poor coverage rates ≤ 7.2% for both exposure metrics (with both analytical and bootstrap-based CIs). In contrast, for SIMEX estimates, bootstrap-based CIs yielded much better coverage rates of 79.2% for current use and 89.8% for cumulative drug use duration in the past 12 weeks, for which SIMEX estimates were less biased.

4 Application: cumulative duration of hydrochlorothiazide use and risk of non-melanoma skin cancer

4.1 Methods

Background

To illustrate a real-life pharmacoepidemiology application, we used a large population-based cohort to re-assess the potential association between TVE to hydrochlorothiazide (HCTZ), a popular antihypertensive drug, and the hazard of non-melanoma skin cancer (NMSC), the most commonly diagnosed cancer (Lomas, Leonardi-Bee and Bath-Hextall, 2012). HCTZ increases the sensitivity of the skin to sunlight and ultraviolet radiation (Blakely, Drucker and Rosen, 2019), which are important risk factors for NMSC (Kaae et al., 2010; Makhzoumi and Arron, 2013). Because the NMSC risk, similar to other cancers, is likely affected by *cumulative* exposure to relevant carcinogen(s) (Danieli et al., 2019; Richardson and Ashmore, 2005), careful modelling of the TVE is required (Abrahamowicz et al., 2012). However, NMSC occurrence can be diagnosed only at a visit to a physician who had requested a biopsy for NMSC or has a relevant specialty, particularly dermatology. Thus, the NMSC occurrence is an interval-censored event.

Data source

We used a subset of participants of the Ontario Health Study (OHS), a long-term study of about 225,000 residents of the Canadian province of Ontario. OHS data are linked to data on all hospitalizations and medical visits in 2006-2017, and provide details on all drug prescriptions for elderly participants, aged over 65 years, covered by the public prescription drug plan.

We relied on the new user design (Ray, 2005), i.e. limited the analyses to subjects who started HCTZ use at some time after they entered the OHS cohort and had *not* used it for six prior months. This design reduces substantially concerns about potential confounding-by-indication related to unmeasured time-invariant characteristics (Ray, 2005). We excluded subjects diagnosed with NMSC or melanoma before their 1st HCTZ prescription. Finally, similar to other studies of this association, we assumed a lag of 2 years (Pottegard et al., 2019; Eworuke et al., 2021) between HCTZ exposure and NMSC occurrence. Accordingly, exposures were lagged (Danieli et al., 2019) and subjects followed for less than 2 years after their 1st HCTZ prescription were excluded, restricting the analyses to 3152 new HCTZ users.

Event and censoring times

Time 0 corresponded to entry into the OHS cohort. To avoid immortal time bias (Suissa, 2008; Zhou et al., 2005), we used the delayed entry approach, with subjects entering risk sets only after the time of their 1st HCTZ prescription (Sylvestre and Abrahamowicz, 2009).

During follow-up, 175 subjects were diagnosed with NMSC. All NMSC diagnoses were recorded in the database at the time of a visit to a physician with a relevant specialty (mostly dermatology, general practice, plastic or general surgery). To account for interval-censored events, all subjects who had no NMSC diagnosis during follow-up were censored at the time of their last recorded relevant physician visit, after which there was *no* opportunity to get the diagnosis. In addition, subjects who discontinued HCTZ treatment or switched to an alternative antihypertension drug were censored 25 months (to account for the 2-year latency discussed above) after the end of their most recent HCTZ prescription. If subject *s* had NMSC diagnosed at a physician visit at time $t_{s,e}$, we imputed the presumed event time at the mid-point of interval between the preceding relevant visit $t_{s,e-1}$ and $t_{s,e}$. Section C.1 of the Supporting Information describes the algorithm used to determine $t_{s,e-1}$, based on the visit history and physician specialties.

Time-varying exposures

Because, for many cancers, the time window of the etiologically relevant exposures is limited to a few years (Danieli et al., 2019; Richardson and Ashmore, 2005), and the median follow-in our cohort was only 4.8 years, we focused on the potential impact of cumulative duration of HCTZ exposure in a 2-year interval spanning 2-4 previous years (to account for the 2-year lag). Section C.2 of the Supporting Information describes details of constructing the resulting continuous TVE.

Analyses

We first estimated the conventional multivariable Cox proportional hazards (PH) model. All models adjusted for calendar year and the following time-invariant covariates, selected *a priori* as known or plausible risk factors for NMSC, evaluated at cohort entry: age group, sex, Charlson comorbidity index, other comorbidities and clinical conditions (diabetes, dialysis, chronic kidney disease, and aid), use of any drug with potential photosensitizing properties, and other antihypertensive drugs. The PH assumption was tested using the Grambsch and Therneau (1994) approach implemented in the R survival package and was *not* rejected either globally (p = 0.084) or for the HCTZ exposure (p = 0.919).

Then, we employed the SIMEX adaptation, proposed in section 2.2, to correct for the potential bias due to interval-censoring of NMSC events. Section C.3 of the Supporting Information provides details of our SIMEX implementation. The confidence interval for SIMEX-corrected log HR was estimated using bootstrap with 1000 resamples.

4.2 Application results

During the total follow-up of 14 310 person-years (mean = 4.5 years, median = 4.8, IQR: 3.2 – 5.8), 175 (5.6%) among the 3152 new HCTZ users were diagnosed with NMSC, with an incidence rate of 12.2 NMSC cases per 1000 person-years.

Table 3 shows that in the conventional multivariable model, increasing duration of exposure, in the 2-year window from 2 to 4 years ago, is associated with a moderate increase in the hazard of NMSC (49% for each additional 2 years of HCTZ exposure). However, this association is statistically non-significant (two-tailed p-value for the model-based Wald test = 0.133), which may be partly due to the low number of 175 events and, thus, limited power.

On the other hand, across the 175 NMSC cases, the mean time difference between the visit when NMSC diagnosis was recorded and the previous relevant medical visit was 8.6 months (median = 5.3, IQR: 3.0 – 9.5). Thus, given simulation results in section 3.2, the above conventional estimate is likely somewhat biased toward the null. Indeed, diagnostics proposed in section 2.3 indicated that the correlation $r(p) = r\{X_s(\tau_s^* - p \cdot \delta t), X_s(\tau_s^*)\}$ between the corresponding exposure values, across the cases, decreased in a fairly *linear* fashion with increased time difference $p \cdot \delta t$, over a long time window of up to 20 months, and tended to stabilize at around r = 0.4 for larger differences (Figure 3). On the other hand, because of the 24-month lag, the observed number of events decreased quickly with artificially increased differences between the two relevant visits for cases, which gradually shifted the imputed event times to earlier dates (see section C.4 and Figure C.1 of the Supporting Information for further comments on this issue). Based on this evidence and recommendations in section 2.3, we decided to (i) use the *linear* extrapolating function at steps 3 and 4 of our proposed SIMEX-like procedure, and (ii) limit the number of iterations at step 1 to k=7, so that the upper threshold for the mean time between relevant visits $\overline{\Delta t}_k$ for cases does not exceed 22.4 months, the last iteration at which the number of observed events (145) remains above 80% of the original 175 events. This corresponds to 5.6 months (22.4/4) in Figure 3.

The resulting SIMEX-corrected estimate suggest a somewhat stronger association than estimated through the conventional model (Table 3). Indeed, the adjusted HRs associated with 2 years of cumulative duration of HCTZ exposure between 2 and 4 years ago, are HR=1.49 (95% CI: 0.89 to 2.52) for conventional model *versus* HR=1.63 (95% CI: 0.92 to 3.15) for the SIMEX-corrected estimate. This corresponds to a 27% relative increase in the estimated *excess* hazard (0.626/0.494 = 1.268), associated with 2 years of continuous HCTZ use in the relevant period of 2-4 years ago after correcting for imprecise timing of the events.

5 Discussion

It is important to develop and validate methods able to handle analytical challenges of assessing associations between time-varying exposures (TVEs) with interval-censored endpoints. We presented a pragmatic solution that adapts the simulation-extrapolation (SIMEX) methodology (Cook and Stefanski, 1994) to time-to-event analyses involving interval-censored events, which occurrence can be established only at sparse clinic visits. We also proposed diagnostics that help choose the SIMEX extrapolating function, used to extrapolate the estimated regression coefficient to the hypothetical ideal situation when event times are measured without errors.

Our simulations, designed to mimic pharmacoepidemiology cohort studies of the association between time-varying drug exposures and the hazard of an interval-censored adverse event, illustrate the potential advantages of the proposed approach. For both a binary indicator of current drug use and cumulative duration of past exposure, SIMEX-corrected estimates reduced considerably the serious bias toward the null of conventional Cox PH model estimates. Even if SIMEX estimates had systematically higher variance, typical for multiple-stage estimation procedures designed primarily to reduce bias (Hernan, Brumback and Robins, 2000; Xiao, Moodie and Abrahamowiczm 2013; Brookhart et al., 2006; Ionescu-Ittu, Delaney and Abrahamowicz, 2009; Esteve et al., 1990; Le Teuff e al., 2005), they generally yielded better bias/variance trade-off, i.e. lower root mean squared error (RMSE) than the corresponding conventional estimates. RMSE reductions were more pronounced for (i) larger samples and/or (ii) stronger true exposure effects, as they imply that, respectively, (i) the variance becomes relatively less important, and (ii) the absolute amount of bias increases (for the same relative bias). Still, the RMSE was generally improved even for N=750 (about 239 events) and weaker exposure effects. As expected, when the true effect was null (HR=1.0), both the conventional and SIMEX estimates were unbiased and similar, except for increased variance of the latter.

Thus, the proposed method may be of particular interest for pharmacoepidemiology which typically relies on very large administrative health databases, with detailed time-varying history of prescriptions filled by several thousands of users of a given drug (Patorno et al., 2015; Abrahamowicz and Tamblyn, 2005), i.e. sample sizes larger than considered in our simulations. As many adverse drug events represent interval-censored endpoints, it is important to reduce the resulting bias in the estimated associations, as illustrated by our analyses of the cumulative impact of exposure to hydrochlorothiazide (HCTZ) on the hazard of non-melanoma skin cancer (NMSC). The fact that the SIMEX-based correction increased the strength of this association only moderately may be partly due to (i) the relatively frequent clinic visits by the participants, and (ii) limited within-subject variation over-time in HCTZ use, consistent with generally good adherence to antihypertensive medications (Friedman et al., 2010; Gee et al., 2012). Thus, the updated values of cumulative exposure duration, for the same subject at different times, were often strongly correlated, reducing the impact of imprecise timing of events. Still, SIMEX-corrected estimates suggested that 2 years of cumulative HCTZ use, 2-4 years ago, were associated with HR=1.63, i.e. a 27% higher estimated *excess* hazard than the conventional HR=1.49.

The classic SIMEX methodology (Cook and Stefanski, 1994) has been extended to several more complex settings, including survival analysis with time-invariant predictors (He, Yi and Xiong, 2007; Greene and Cai, 2004; Zhang, He and Li, 2014), estimation of IPT weights in marginal structural models (Kyle et al., 2016), non-linear effects of error-prone time-varying covariates (Wang et al., 2020), or categorical predictors (Kuchenhoff et al., 2006). Similar to our approach, most of these extensions used some *ad hoc* methods and then relied on simulations to assess and validate the resulting SIMEX-corrected estimates (Kyle et al., 2016; Kuchenhoff et al., 2006; Wang et al., 2020; Oh et al., 2018).

Recently, SIMEX was adapted to address a somewhat different problem of inaccurate timing of events self-reported by the patients (Oh et al., 2018). The authors assumed that the resulting discrepancies from the true (unknown) event times follow the classic ME model, with normal distribution around the mean of 0. Their simulation results are generally similar to ours in that SIMEX-corrected estimates considerably reduce bias of the conventional Cox model-based estimates, but do *not* entirely eliminate bias. However, TVEs are *not* discussed while, based on our simulation results, they may be especially sensitive to inaccuracies in the event times. Thus, further research should evaluate SIMEX-corrected estimates of the association between time-varying exposures and the hazard of events for which times are affected by classic measurement errors.

We recognize some limitations of both our methods and our simulations. Firstly, our pragmatic SIMEX-like approach relies on the choice of an appropriate functional form of the extrapolating function used in steps 3 and 4. Indeed, for

the binary time-varying indicator of current drug use, for example, both the bias and RMSE of SIMEX-corrected estimates may be only marginally lower than for conventional estimates if the extrapolating function is *not* well chosen (Tables B.2 and B.4 in the Supporting Information). To address this practically important challenge, in section 2.3 we propose specific diagnostics that rely on the fact that in many real-life applications, e.g. pharmacoepidemiology analyses of prescription databases or environmental studies of ambient air pollution, the full history TVE is recorded. Therefore, the expected impact of (a) increasing absolute discrepancy between imputed vs. true event times on (b) decreasing correlation between the corresponding updated exposure values can be approximated based on the available data. One can rely on plots of the relationship between (a) and (b) to choose the analytical form of the extrapolating function. Yet, simulation results in section 3.2 suggest using linear function as a default option, unless the plot shows considerable non-linearity. The correlation plot may also often suggest an approximate interval over which the difference between visits is artificially increased and, thus, the number of iterations for SIMEX step 1. However, as illustrated in real-life application in section 4, if the exposure is lagged, the number of iterations may have to be reduced to avoid losing too many early events (section C.4 of the Supporting Information provide further insights). Finally, it is possible that in some applications the proposed diagnostics will be of little help. For example, if the average difference D between times of the two relevant visits, before and after the true event time, is so large that the correlation between the corresponding exposure values is close to 0, the proposed extrapolation will *not* work due to a "floor effect", yielding the SIMEX-based extrapolated log HR value very close to the conventional uncorrected estimate. If so, the proposed diagnostic plot may be still help approximately assess whether the resulting estimates are highly or only moderately biased, but the correction will not be possible. In addition, while we focused on MEs in interval-censored event times, similar errors may affect time-varying covariates, e.g. the occurrence of new comorbidities, based on clinical assessment at discrete times of medical visits. This may induce residual confounding of the estimated exposure effect, with the direction of bias that depends on the pattern of covariate-exposure and covariate-outcome associations (Brenner and Blettner, 1997). Further analytical developments will be necessary to allow SIMEX-like corrections for errors in both interval-censored events and time-varying covariates.

Furthermore, the potential advantages of the proposed SIMEX-based approach depend on both the effective sample size, i.e. the number of uncensored events, and the underlying data structure. From this perspective, it is important to ensure that simulations that assess the performance of statistical methods are based on realistic assumptions, relevant for potential future real-life applications (Boulesteix et al., 2018; Morris et al., 2019). Accordingly, our time-varying simulations were designed to generate plausible patterns of both between- and within-patient changes in time-varying exposure. Future research may consider plasmode simulations, with exposure matrix based on specific real-life dataset (Franklin et al., 2014).

Finally, similar to other SIMEX implementations, the multi-step approach proposed in section 2.2 requires repeated iterations of data manipulations and refitting of multivariable models. However, necessary calculations are rather straightforward and the resulting computing times are reasonable. For example, the time to get the SIMEX estimate for the current use metric for one simulated sample with N=3000 for scenario 1 was about 45 seconds and 3.5 hours, respectively, without and with calculating the 95% bootstrap CI based on 300 resamples (computer under Windows 10 operating system, with Intel Xeon 3.60 GHz processor).

In conclusion, we proposed a pragmatic extension of the SIMEX methodology to correct for bias in time-to-event analyses of associations between time-varying exposures and the hazard of interval-censored events. Given encouraging results of both simulations and the real-life pharmacoepidemiology application, we believe this method may be useful in several large empirical studies. Our work may also stimulate further methodological research on adapting SIMEX to deal with different challenges specific to time-to-event analyses.

Acknowledgements This research was supported by the Natural Sciences and Engineering Research Council of Canada (grant 228203, Principal Investigator (PI) M. Abrahamowicz) and the Canadian Institutes of Health Research (grant PJT-148946, PI M. Abrahamowicz; grant TD3-137716, Nominated PI S. Bernatsky). Michal Abrahamowicz is a James McGill Professor at McGill University. The authors thank Dr. Christine Binquet for insightful comments on the manuscript.

Conflict of Interest

The authors have declared no conflict of interest.

Data availability statement

Simulation data that support the findings of this study can be generated with the code provided as Supporting Information. Real-

life data of the application presented cannot be shared due to privacy and legal restrictions.

References

Abrahamowicz, M., Beauchamp, M. E., and Sylvestre, M. P. (2012). Comparison of alternative models for linking drug exposure with adverse effects. *Statistics in Medicine*, 31, 1014-1030.

Abrahamowicz, M., and Tamblyn, R. (2005). Drug utilization patterns. In *Encyclopedia of Biostatistics*, eds. P. Armitage and T. Colton, Chichester: John Wiley & Sons, Ltd.

Ahn, S., Lim, J., Paik, M. C., Sacco, R. L., and Elkind, M. S. (2018). Cox model with interval-censored covariate in cohort studies. *Biometrical Journal*, 60, 797-814.

Andersen, P. K., and Liestol, K. (2003). Attenuation caused by infrequently updated covariates in survival analysis. *Biostatistics*, 4, 633-649.

Andersen, P. K., Pohar Perme, M., van Houwelingen, H. C., Cook, R. J., Joly, P., Martinussen, T., Taylor, J. M. G., Abrahamowicz, M., and Therneau, T. M. (2021). Analysis of time-to-event for observational studies: Guidance to the use of intensity models. *Statistics in Medicine*, 40, 185-211.

Avorn, J. (2007). In defense of pharmacoepidemiology - Embracing the yin and yan of drug research. *New England Journal of Medicine*, 357, 2219-2221.

Bacchetti, P., and Quale, C. (2002). Generalized additive models with interval-censored data and time-varying covariates: application to human immunodeficiency virus infection in hemophiliacs. *Biometrics*, 58, 443-447.

Blakely, K. M., Drucker, A. M., and Rosen, C. F. (2019). Drug-induced photosensitivity - an update: culprit drugs, prevention and management. *Drug Safety*, 42, 827-847.

Boulesteix, A. L., Binder, H., Abrahamowicz, M., and Sauerbrei, W. (2018). On the necessity and design of studies comparing statistical methods. *Biometrical Journal*, 60, 216-218.

Brenner, H., and Blettner, M. (1997). Controlling for continuous confounders in epidemiologic research. *Epidemiology*, 8, 429-434.

Brookhart, M. A., Wang, P. S., Solomon, D. H., and Schneeweiss, S. (2006). Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable. *Epidemiology*, 17, 268-275.

Campbell, K. R., Juarez-Colunga, E., Grunwald, G. K., Cooper, J., Davis, S., and Gralla, J. (2019). Comparison of a time-varying covariate model and a joint model of time-to-event outcomes in the presence of measurement error and interval censoring: application to kidney transplantation. *BMC Medical Research Methodology*, 19, 130.

Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006). *Measurement error in nonlinear models: a modern perspective.* CRC Press.

Chen, E. B., and Cook, R. J. (2003). Regression modeling with recurrent events and time-dependent interval-censored marker data. *Lifetime Data Analysis*, 9, 275-291.

Chiang, Y. K., Hardy, R. J., Hawkins, C. M., and Kapadia, A. S. (1989). An illness-death process with time-dependent covariates. *Biometrics*, 45, 669-681.

Clark, D. E., Doolittle, P. C., Winchell, R. J., and Betensky, R. A. (2014). The effect of hospital care on early survival after penetrating trauma. *Injury Epidemiology*, 1, 24.

Clark, D. E., Winchell, R. J., and Betensky, R. A. (2013). Estimating the effect of emergency care on early survival

after traffic crashes. *Accident Analysis & Prevention*, 60, 141-147.

Cook, J. R., and Stefanski, L. A. (1994). Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association*, 89, 1314-1328.

Cox, D. R. (1972) Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B*, 34, 187-220.

Danieli, C., Cohen, S., Liu, A., Pilote, L., Guo, L., Beauchamp, M. E., Marelli, A. J., and Abrahamowicz, M. (2019). Flexible Modeling of the Association Between Cumulative Exposure to Low-Dose Ionizing Radiation From Cardiac Procedures and Risk of Cancer in Adults With Congenital Heart Disease. *American Journal of Epidemiology*, 188, 1552-1562.

Esteve, J., Benhamou, E., Croasdale, M., and Raymond, L. (1990). Relative survival and the estimation of net survival: elements for further discussion. *Statistics in Medicine*, 9, 529-538.

Eworuke, E., Haug, N., Bradley, M., Cosgrove, A., Zhang, T., Dee, E. C., Adimadhyam, S., Petrone, A., Lee, H., Woodworth, T., and Toh, S. (2021). Risk of Nonmelanoma Skin Cancer in Association With Use of Hydrochlorothiazide-Containing Products in the United States. *JNCI Cancer Spectrum*, 5, kab009.

Finkelstein, D. M. (1986). A proportional hazards model for interval-censored failure time data. *Biometrics*, 42, 845-854.

Franklin, J. M., Schneeweiss, S., Polinski, J. M., and Rassen, J. A. (2014). Plasmode simulation for the evaluation of pharmacoepidemiologic methods in complex healthcare databases. *Computational Statistics and Data Analysis*, 72, 219-226.

Friedman, O., McAlister, F. A., Yun, L., Campbell, N. R., and Tu, K. (2010). Antihypertensive drug persistence and compliance among newly treated elderly hypertensives in ontario. *American Journal of Medicine*, 123, 173-181.

Gee, M. E., Campbell, N. R., Gwadry-Sridhar, F., Nolan, R. P., Kaczorowski, J., Bienek, A., Robitaille, C., Joffres, M., Dai, S., and Walker, R. L. (2012). Antihypertensive medication use, adherence, stops, and starts in Canadians with hypertension. *Canadian Journal of Cardiology*, 28, 383-389.

Gentleman, R., and Geyer, C. J. (1994). Maximum likelihood for interval censored data: Consistency and computation. *Biometrika*, 81, 618-623.

Grambsch, P. M., and Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrics*, 81, 515-526.

Greene, W. F., and Cai, J. (2004). Measurement error in covariates in the marginal hazards model for multivariate failure time data. *Biometrics*, 60, 987-996.

He, W., Yi, G. Y., and Xiong, J. (2007). Accelerated failure time models with covariates subject to measurement error. *Statistics in Medicine*, 26, 4817-4832.

Hernan, M. A., Brumback, B., and Robins, J. M. (2000). Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology*, 11, 561-570.

Ionescu-Ittu, R., Delaney, J. A., and Abrahamowicz, M. (2009). Bias-variance trade-off in pharmacoepidemiological studies using physician-preference-based instrumental variables: a simulation study. *Pharmacoepidemiology and Drug Safety*, 18, 562-571.

Kaae, J., Boyd, H. A., Hansen, A. V., Wulf, H. C., Wohlfahrt, J., and Melbye, M. (2010). Photosensitizing medication use and risk of skin cancer. *Cancer Epidemiology, Biomarkers & Prevention*, 19, 2942-2949.

Kooperberg, C., and Clarkson, D. B. (1997). Hazard regression with interval-censored data. *Biometrics*, 53, 1485-1494.

Kooperberg, C., and Stone, C. J. (1992). Logspline density estimation for censored data. *Journal of Computational and Graphical Statistics*, 1, 301-328.

Kuchenhoff, H., Mwalili, S. M., and Lesaffre, E. (2006). A general method for dealing with misclassification in regression: the misclassification SIMEX. *Biometrics*, 62, 85-96.

Kyle, R. P., Moodie, E. E. M., and Abrahamowicz, M. (2016). Correcting for measurement error in time-varying covariates in marginal structural models. *American Journal of Epidemiology*, 184, 249-258.

Le Teuff, G., Abrahamowicz, M., Bolard, P., and Quantin, C. (2005) Comparison of Cox's and relative survival models when estimating the effects of prognostic factors on disease-specific mortality: a simulation study under proportional excess hazards. *Statistics in Medicine*, 24, 3887-3909.

Lindsey, J. C., and Ryan, L. M. (1998). Tutorial in biostatistics methods for interval-censored data. *Statistics in Medicine*, 17, 219-238.

Lomas, A., Leonardi-Bee, J., and Bath-Hextall, F. (2012). A systematic review of worldwide incidence of nonmelanoma skin cancer. *British Journal of Dermatology*, 166, 1069-1080.

MacMahon, S., and Collins, R. (2001). Reliable assessment of the effects of treatment on mortality and major morbidity, II: observational studies. *Lancet*, 357 , 455-462.

Makhzoumi, Z. H., and Arron, S. T. (2013). Photosensitizing agents and the risk of non-melanoma skin cancer: a population-based case-control study. *Journal of Investigative Dermatology*, 133, 1922-1923.

Morris, T. P., White, I. R., and Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine*, 38, 2074-2102.

Oh, E. J., Shepherd, B. E., Lumley, T., and Shaw, P. A. (2018) Considerations for analysis of time-to-event outcomes measured with error: Bias and correction with SIMEX. *Statistics in Medicine*, 37, 1276-1289.

Oller, R., and Gomez, M. G. (2020). A nonparametric test for the association between longitudinal covariates and censored survival data. *Biostatistics*, 21, 727-742.

Patorno, E., Garry, E. M., Patrick, A. R., Schneeweiss, S., Gillet, V. G., Zorina, O., Bartels, D. B., and Seeger, J. D. (2015). Addressing limitations in observational studies of the association between glucose-lowering medications and all-cause mortality: a review. *Drug Safety*, 38, 295-310.

Pazzagli, L., Linder, M., Zhang, M., Vago, E., Stang, P., Myers, D., Andersen, M., and Bahmanyar, S. (2018). Methods for time-varying exposure related problems in pharmacoepidemiology: An overview. *Pharmacoepidemiology and Drug Safety*, 27, 148-160.

Pottegard, A., Pedersen, S. A., Schmidt, S. A. J., Lee, C. N., Hsu, C. K., Liao, T. C., Shao, S. C., and Lai, E. C. (2019) Use of hydrochlorothiazide and risk of skin cancer: a nationwide Taiwanese case-control study. *British Journal of Cancer*, 121, 973-978.

Ray, W. A. (2005). Observational studies of drugs and mortality. *New England Journal of Medicine*, 353, 2319-2321.

Richardson, D. B., and Ashmore, J. P. (2005). Investigating time patterns of variation in radiation cancer associations. *Occupational and Environmental Medicine*, 62, 551-558.

Schoenfeld, D. A., Rajicic, N., Ficociello, L. H., and Finkelstein, D. M. (2011). A test for the relationship between a

time-varying marker and both recovery and progression with missing data. *Statistics in Medicine*, 30, 718-724.

Seaman, S. R., and Bird, S. M. (2001). Proportional hazards model for interval-censored failure times and time-dependent covariates: application to hazard of HIV infection of injecting drug users in prison. *Statistics in Medicine*, 20, 1855-1870.

Sparling, Y. H., Younes, N., Lachin, J. M., and Bautista, O. M. (2006). Parametric survival models for interval-censored data with time-dependent covariates. *Biostatistics*, 7, 599-614.

Stefanski, L. A., and Cook, J. R. (1995). Simulation extrapolation: the measurement error jackknife. *Journal of the American Statistical Association*, 90 , 1247-1256.

Suissa, S. (2008). Immortal time bias in pharmacoepidemiology. *American Journal of Epidemiology*, 167, 492-499.

Sun, X., Li, X., Chen, C., and Song, Y. (2013). A review of statistical issues with progression-free survival as an interval-censored time-to-event endpoint. *Journal of Biopharmaceutical Statistics*, 23, 986-1003.

Sylvestre, M. P., and Abrahamowicz, M. (2008) Comparison of algorithms to generate event times conditional on time-dependent covariates. *Statistics in Medicine*, 27, 2618-2634.

Sylvestre, M. P. and Abrahamowicz, M. (2009). Flexible modeling of the cumulative effects of time-dependent exposures on the hazard. *Statistics in Medicine,* 28, 3437-3453.

Sylvestre, M. P., Evans, T., MacKenzie, T., and Abrahamowicz, M. (2015). PermAlgo: Permutational algorithm to generate event times conditional on a covariate matrix including time-dependent covariates. R package, version 1.1. Retrieved from https://CRAN.R-project.org/package=PermAlgo.

Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society, Series B*, 38, 290-295.

Wang, Y., Beauchamp, M. E., and Abrahamowicz, M. (2020). Nonlinear and time-dependent effects of sparsely measured continuous time-varying covariates in time-to-event analysis. *Biometrical Journal*, 62, 492-515.

Xiao, Y., Moodie, E. E. M., and Abrahamowicz, M. (2013). Comparison of approaches to weight truncation for marginal structural Cox models. *Epidemiologic Methods*, 2, 1-20.

Zhan, Z., de Bock, G. H., Wiggers, T., and van den Heuvel, E. (2016) The analysis of terminal endpoint events in stepped wedge designs. *Statistics in Medicine*, 35, 4413-4426.

Zhang, J., He, W., and Li, H. (2014). A semiparametric approach for accelerated failure time models with covariates subject to measurement error. *Communications in Statistics - Simulation and Computation*, 43, 329-341.

Zhang, Z., and Sun, J. (2010). Interval censoring. *Statistical Methods in Medical Research*, 19, 53-70.

Zhou, Z., Rahme, E., Abrahamowicz, M., and Pilote, L. (2005). Survival bias associated with time-to-treatment initiation in drug effectiveness evaluation: a comparison of methods. *American Journal of Epidemiology*, 162, 1016-1023.

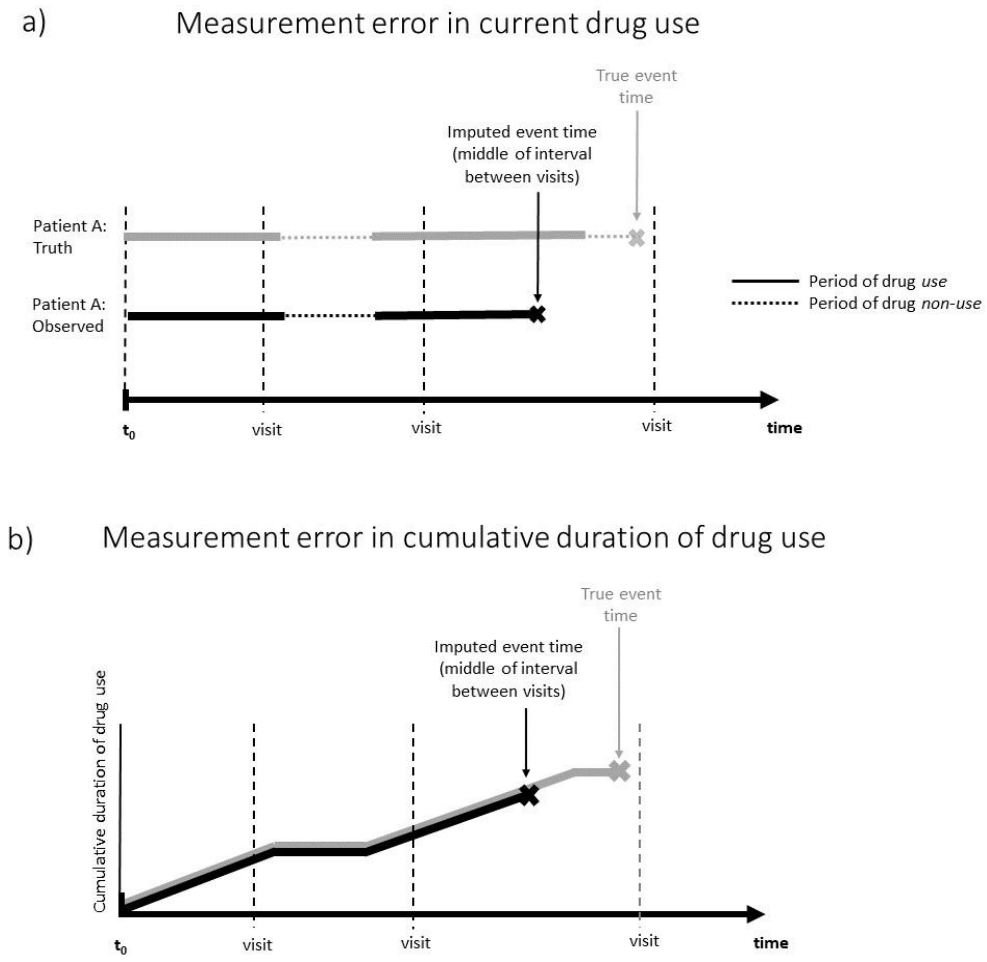

Figure 1 Measurement errors in time-varying exposure metrics due to imputing the event time at the mid-point between the two relevant visits, relative to exposure metric values at the true event time, for a hypothetical subject. The two panels correspond to the exposure metrics used in our simulations: a) current drug use $U(t)$, and b) cumulative duration of drug use in the past 12 weeks $Cum12(t)$. In panel a), the true event time occurs during a period of non-use of the drug, but the imputed event time at the mid-point of the interval between the two relevant visits correspond to a period of current use. In panel b) it can be seen that the cumulative duration of drug use differs between the imputed vs. true event times.

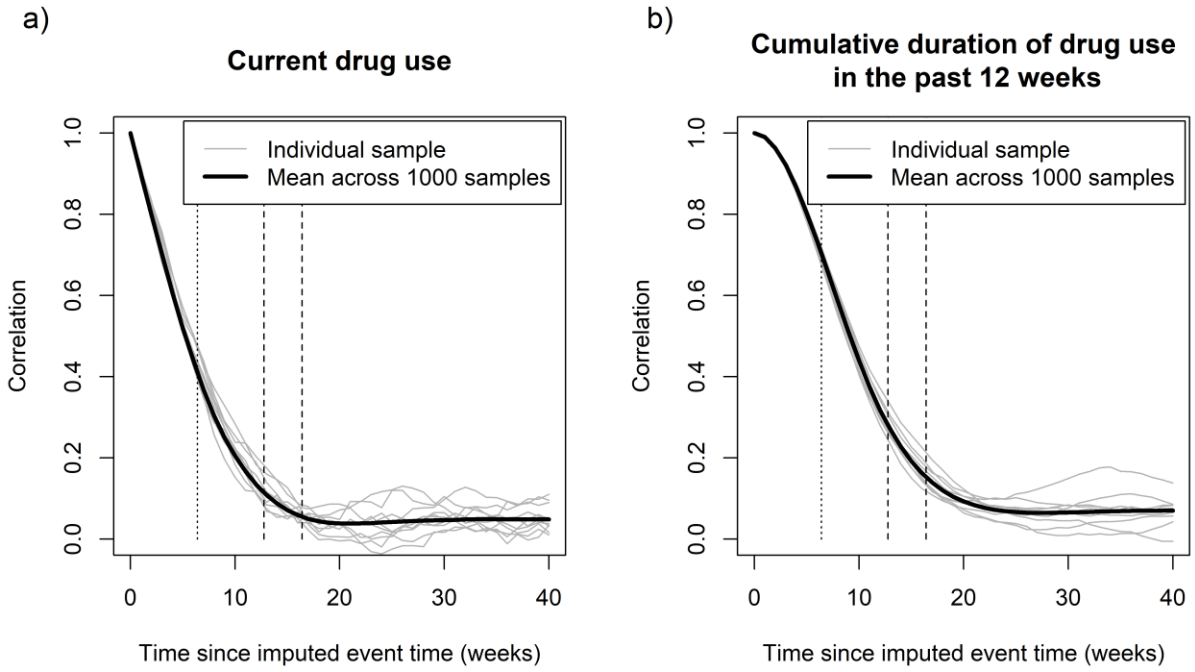

Figure 2 Diagnostics for the selection of the extrapolating function for each of the two exposure metrics: a) current drug use $U(t)$ and b) cumulative duration of drug use in the past 12 weeks $Cum12(t)$, for scenario 1 (N=3000, stronger exposure effects, and duration-based exposure generation). Each panel shows how the correlation between the exposure values observed for the same case at different times varies with increasing time difference from the original imputed event time. The gray lines are for 10 individual samples and the bold black line represents the mean across the 1000 simulated samples. The vertical dotted line indicates the correlation corresponding to the mean difference between the imputed vs. the true event times (at 25.6/4 = 6.4 weeks, given observed $\overline{\Delta t}$= 25.6 weeks on average across simulated samples). The vertical dashed lines in each panel correspond to the selected number of iterations for the proposed SIMEX approach, i.e. either 15 or 25 iterations.

Table 1 Comparison of model-specific log HRs for the main scenarios with duration-based exposure generation

Scenario	*N* [mean # events][1]	True log HR	Cox model	SIMEX extrapolating $f(\cdot)$	Relative bias (%)	SD	RMSE
Exposure metric: Current use							
1	3000 [956]	log(3) = 1.10	Conventional, event at middle of interval	--	-50.4	0.065	0.558
			SIMEX	Linear 15 iterations	-19.7	0.140	0.257
			SIMEX	Linear 25 iterations	-24.5	0.114	0.292
2	3000 [955]	log(2) = 0.69	Conventional, event at middle of interval	--	-49.1	0.062	0.346
			SIMEX	Linear 15 iterations	-16.2	0.142	0.181
			SIMEX	Linear 25 iterations	-21.4	0.110	0.185
3	1500 [478]	log(3) = 1.10	Conventional, event at middle of interval	--	-50.7	0.092	0.564
			SIMEX	Linear 15 iterations	-19.8	0.201	0.296
			SIMEX	Linear 25 iterations	-24.2	0.160	0.310
4	1500 [478]	log(2) = 0.69	Conventional, event at middle of interval	--	-49.0	0.090	0.351
			SIMEX	Linear 15 iterations	-17.7	0.195	0.231
			SIMEX	Linear 25 iterations	-21.4	0.157	0.216
5	750 [240]	log(3) = 1.10	Conventional, event at middle of interval	--	-50.2	0.127	0.566
			SIMEX	Linear 15 iterations	-19.6	0.281	0.354
			SIMEX	Linear 25 iterations	-24.1	0.224	0.347
6	750 [239]	log(2) = 0.69	Conventional, event at middle of interval	--	-49.4	0.131	0.367
			SIMEX	Linear 15 iterations	-18.8	0.300	0.327
			SIMEX	Linear 25 iterations	-22.5	0.236	0.283
Exposure metric: Cumulative duration in past 12 weeks							
1	3000 [956]	log(2)/6 = 0.12	Conventional, event at middle of interval	--	-30.7	0.007	0.036
			SIMEX	Linear 15 iterations	8.3	0.015	0.018
			SIMEX	Linear 25 iterations	6.4	0.014	0.015
2	3000 [955]	log(3)/12 = 0.09	Conventional, event at middle of interval	--	-30.4	0.006	0.029
			SIMEX	Linear 15 iterations	8.0	0.014	0.016
			SIMEX	Linear 25 iterations	6.4	0.012	0.014
3	1500 [479]	log(2)/6 = 0.12	Conventional, event at middle of interval	--	-30.6	0.010	0.037
			SIMEX	Linear 15 iterations	8.3	0.022	0.024
			SIMEX	Linear 25 iterations	6.0	0.019	0.020

4	1500 [478]	log(3)/12 = 0.09	Conventional, event at middle of interval	--	-30.2	0.010	0.029
			SIMEX	Linear 15 iterations	8.5	0.020	0.022
			SIMEX	Linear 25 iterations	7.3	0.018	0.019
5	750 [240]	log(2)/6 = 0.12	Conventional, event at middle of interval	--	-29.8	0.013	0.037
			SIMEX	Linear 15 iterations	9.7	0.030	0.032
			SIMEX	Linear 25 iterations	7.4	0.026	0.027
6	750 [239]	log(3)/12 = 0.09	Conventional, event at middle of interval	--	-30.4	0.014	0.031
			SIMEX	Linear 15 iterations	7.7	0.029	0.029
			SIMEX	Linear 25 iterations	6.3	0.025	0.026

[1] Mean number of events in imputed data. Events after the last visit during follow-up are ignored, as subjects are censored at their last visit.

Table 2 Comparison of model-specific log HRs for additional scenarios with duration-based exposure generation

Scenario	***N* [mean # events][1]**	**True log HR**	**Cox model**	**SIMEX extrapolating $f(\cdot)$**	**Relative bias (%)**	**SD**	**RMSE**
Exposure metric: Current use							
7 Extra TV risk factor	3000 [955]	log(3) = 1.10	Conventional, event at middle of interval	--	-50.6	0.068	0.561
			SIMEX	Linear 15 iterations	-19.5	0.149	0.261
			SIMEX	Linear 25 iterations	-24.6	0.118	0.295
8	3000 [955]	log(1) = 0	Conventional, event at middle of interval	--	bias: -0.001	0.064	0.064
			SIMEX	Linear 15 iterations	bias: 0.003	0.147	0.147
			SIMEX	Linear 25 iterations	bias: 0.002	0.114	0.114
Exposure metric: Cumulative duration in past 12 weeks							
7 Extra TV risk factor	3000 [955]	log(2)/6 = 0.12	Conventional, event at middle of interval	--	-30.8	0.007	0.036
			SIMEX	Linear 15 iterations	8.1	0.015	0.017
			SIMEX	Linear 25 iterations	6.5	0.013	0.015
8	3000 [955]	log(1) = 0	Conventional, event at middle of interval	--	bias: 0.000	0.007	0.007
			SIMEX	Linear 15 iterations	bias: 0.000	0.014	0.014
			SIMEX	Linear 25 iterations	bias: 0.000	0.013	0.013

[1] Mean number of events in imputed data. Events after the last visit during follow-up are ignored, as the subjects are censored at the last visit.

Table 3 Results of alternative multivariable Cox PH models for the association of cumulative hydrochlorothiazide (HCTZ) use duration with the hazard of non-melanoma skin cancer

Exposure	Cox model	Adjusted HR for 2-year increase	95% CI
HCTZ duration of use, in the 2-year window from 4 to 2 years ago	Conventional, event at middle of interval	1.49	(0.89, 2.52)[1]
	SIMEX, linear extrapolation 7 iterations	1.63	(0.92, 3.15)[2]

[1] Covariance matrix-based confidence interval (CI).
[2] CI based on 1000 bootstrap resamples.

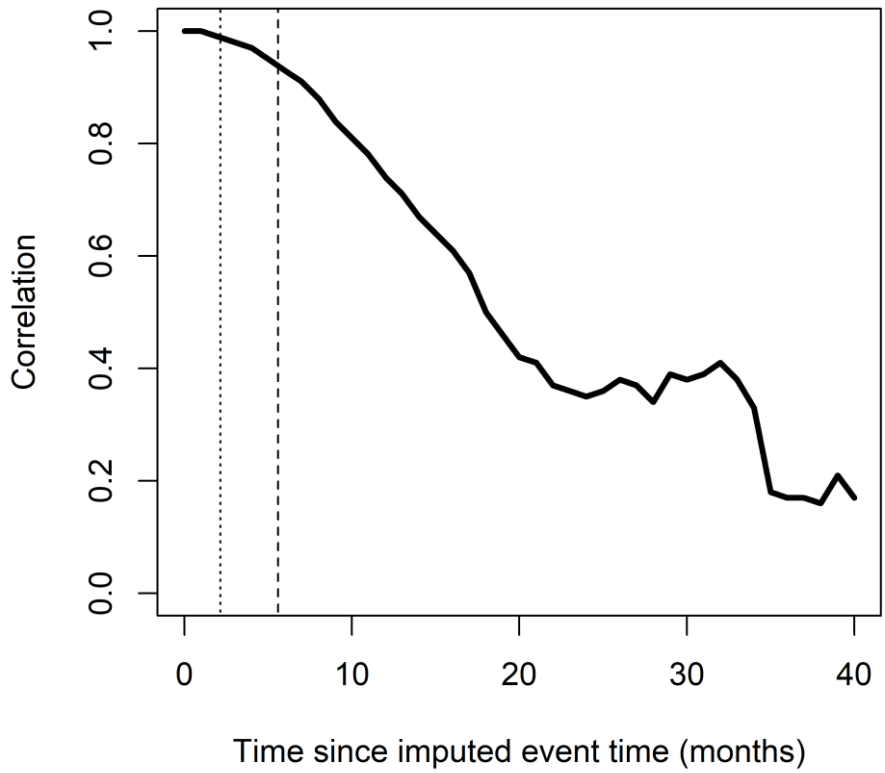

Figure 3 Diagnostics for the selection of the extrapolating function form, for the association of cumulative 2-year duration, 2-4 years ago, of hydrochlorothiazide use and the risk of non-melanoma skin cancer. The graph shows how the correlation between the exposure values observed for the same case at different times varies with increasing time difference from the original imputed event time. The vertical dotted line indicates the correlation corresponding to the mean difference between the imputed vs. the true event times in the observed data (8.6/4 = 2.1 months, given the observed $\overline{\Delta t}$ = 8.6 months). The vertical dashed line corresponds to the selected number of iterations for the SIMEX approach, i.e. 7 iterations, corresponding to 5.6 months on x axis.