Statistics
in Medicine

Special Issue Paper

Received 1 December 2010, Accepted 29 June 2011 Published online 17 November 2011 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.4343

Comparison of alternative models for linking drug exposure with adverse effects

Michal Abrahamowicz,[a,b*†] Marie-Eve Beauchamp[b] and Marie-Pierre Sylvestre[c,d]

Pharmacoepidemiology investigates associations between time-varying medication use/dose and risk of adverse events. Applied research typically relies on a priori chosen simple conventional models, such as current dose or any use in the past 3 months. However, different models imply different risk predictions, and only one model can be etiologically correct in any specific applications. We first formally defined several candidate models mapping the time vector of past drug doses ($X(t),\ t = 1,\ldots,u$) into the value of a time-varying exposure metric $M(u)$ at current time u. In addition to conventional one-parameter models, we considered two-parameter models accounting for recent dose increase or withdrawal and a flexible spline-based weighted cumulative exposure (WCE) model that defines $M(u)$ as the weighted sum of past doses. In simulations, we generated event times assuming one of the models was correct and then analyzed the data with all candidate models. We demonstrated that the minimum AIC criterion is able to identify the correct model as the best-fitting model or one of the equivalent (within 4 AIC points of the minimum) models in a vast majority of simulated samples, especially with 500 or more events. We also showed how relying on an incorrect a priori chosen model may largely reduce the power to test for an association. Finally, we demonstrated how the flexible WCE estimates may help with model diagnostics even if the correct model is not WCE. We illustrated the practical advantages of AIC-based a posteriori model selection and WCE modeling in a real-life pharmacoepidemiology example. Copyright © 2011 John Wiley & Sons, Ltd.

Keywords: pharmacoepidemiology; model selection; goodness of fit; time-varying covariates; cumulative effects; power

1. Introduction

Pharmacoepidemiology attempts to detect adverse effects of medications and typically relies on large administrative health databases [1]. Such databases provide details on duration and dosage of all drug prescriptions received by all individuals in a source population [2, 3]. In real-life settings, both the dose and the duration of treatment vary considerably both between subjects and within subject over time. For example, during 5 years, the total duration of use of benzodiazepines (psychotropic drugs) among 70,000 elderly subjects varied from 1 to 1826 days (mean = 229, median = 83, standard deviation (SD) = 327) and the number of distinct periods of exposure ranged from 1 to 32. Daily dose ranged from 0.1 to 11 times the WHO recommended dose (median = 0.48, SD = 0.37). Some subjects increased, whereas others decreased their dose during the follow-up [4].

Linking such longitudinal patterns of time-varying drug dose/use with adverse events poses conceptual and methodological challenges. Usually, there is little prior knowledge regarding how the variation in the past doses $X(t)$, $t = 1,\ldots,u$, affects the current adverse event risk at time u. Recent pharmacoepidemiological studies typically rely on simple ad hoc 'conventional' models such as current

[a]*Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, QC, Canada*
[b]*Division of Clinical Epidemiology, McGill University Health Centre, Montreal, QC, Canada*
[c]*Research Centre of the University of Montreal Hospital Centre, Montreal, QC, Canada*
[d]*Department of Social and Preventive Medicine, University of Montreal, Montreal, QC, Canada*
**Correspondence to: Michal Abrahamowicz, Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, QC, Canada.*
[†]*E-mail: michal.abrahamowicz@mcgill.ca*

use [5], any drug use in the past [6], recent use in an arbitrary-defined time window, length of which may vary across different studies of the *same* drug [7, 8], current [5] or average recent daily dose [7], and total cumulative past dose or duration of past drug exposure [6]. Clearly, each conventional model implies different impact of past drug use/dose on current risks.

Whereas only a *single* model is etiologically correct for an association between a specific drug and a particular adverse event, *different* models may accurately describe associations for different combinations of drugs and adverse events. Such differences may reflect pharmacokinetics/pharmacodynamics of the drug and/or latency of the biological mechanism linking its use with a given clinical endpoint [9]. For example, the *current* use/dose of psychotropic medications may have *acute short-term* effects on the user's cognitive function [10]. In contrast, the association between prednisone, an anti-inflammatory drug, and cardiovascular risk [11] is likely mediated through *gradual* increases in cardiovascular risk factors [12]. Thus, current cardiovascular risks may be mostly affected by *long-term cumulative* prednisone dose in the *past*.

In the absence of firm prior knowledge about the etiologically correct exposure–risk model, researchers may use empirical criteria to compare goodness of fit of alternative models linking past drug exposure patterns with current risks. Indeed, intuitively one expects that the best-fitting model will be relatively close to the (unknown) correct model. Identification of the correct model may both increase the power for detecting the association and help rationalize the prescribing practice, resulting in improved benefits/harms ratio.

In this paper, we first formally define alternative plausible models linking drug use/dose patterns with adverse events, including some relatively novel models. We then illustrate the underlying differences in modeling time-varying risks associated with a given drug exposure pattern. Next, in simulations, we evaluate to what extent statistical goodness-of-fit criteria are able to identify the etiologically correct among several candidate models and illustrate the practical implications of relying on the incorrect model. Finally, we compare the fit of alternative models in a real-life pharmacoepidemiological application.

2. Alternative models for time-varying drug exposure

We describe in the following text selected models for modeling the impact of a time-varying drug exposure on the hazard of an adverse event. Each model defines a model-specific time-dependent covariate $M(u)$ representing the current value of an exposure metric as a function of past doses:

$$M(u) = f[X(1), X(2), \ldots, X(u-1), X(u)] \tag{1}$$

where u indicates the current time when the hazard is assessed and $\mathbf{X}(t)$, $t = 1, \ldots, u$, is the time vector of past doses.

The time-varying exposure metric $M(u)$ maps the vector of past drug doses $\mathbf{X}(t)$, $t = 1, \ldots, u$, on the unidimensional axis of 'etiologically relevant exposure' at time u. In other words, in any given application, $M(u)$ is defined (implicitly or explicitly) as to reflect the investigators expectations, beliefs, and/or previous findings regarding how the current and past use of the drug may be related to the outcome of interest. By definition, if the subject was never exposed until the current time u (implying $\mathbf{X}(t) = 0$, for any $\mathbf{X}(t)$, $t \leqslant u$), then $M(u) = 0$. As discussed in the following text, usually some additional regularity conditions are imposed on $M(u)$, on the basis of substantive knowledge. For example, $M(u)$ may be restricted to have nonnegative values over only a limited time interval, ending at the current day u and beginning at time $u - t$ in the past.

Once $M(u)$ has been calculated for all times u during the follow-up, standard regression methods, such as Cox proportional hazards (PH) model with time-dependent covariates, can be used to estimate and test its association with the hazard. Still, the challenge is to define the etiologically correct function $f[.]$ mapping a time vector of past doses $\mathbf{X}(t)$, $t = 1, \ldots, u$, into a single current value of exposure metric $M(u)$.

2.1. Conventional parametric models

First, we consider some conventional models, selected from a large set of simple ad hoc models used in applied pharmacoepidemiological studies:

Model 1. Current use: $M(u) = I[X(u) > 0]$, where $I[.]$ is a binary indicator function;
Model 2. Current dose: $M(u) = X(u)$;

Model 3. Nonlinear effect of current dose (NL current dose): $M(u) = g[X(u)]$, where $g[.]$ represents a nonlinear, a priori selected, usually parametric transformation of the dose;
Model 4. Total duration of past use (duration past use): $M(u) = \sum(I[X(t) > 0]$, for $0 < t \leqslant u)$;
Model 5. Total cumulative past dose (cumulative past dose): $M(u) = \sum[X(t)$, for $0 < t \leqslant u]$;
Model 6. Total cumulative dose in past m days (cumulative dose past m days): $M(u) = \sum[X(t)$, for $(u - m) < t \leqslant u]$.

To illustrate the implications of the alternative models, we compare in Figure 2 the resulting time-dependent exposure metrics $M(u)$ for the *same* hypothetical subject, whose doses over 365 days of follow-up, $\mathbf{X}(t), t = 1, \ldots, 365$, are shown in Figure 1.

The differences between the model-specific $M(u)$ in Figure 2(a)–(f) reflect the differences in the underlying a priori assumptions about the relative importance of doses (or drug use) at different times in

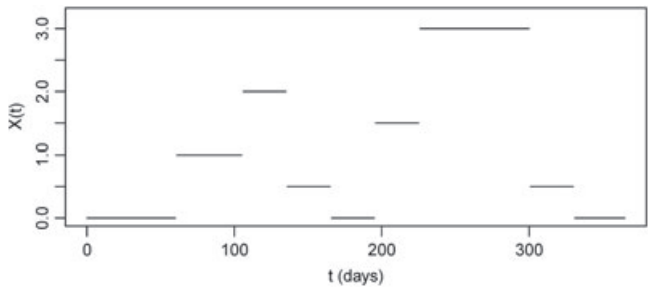

Figure 1. Daily dose *X(t)* for a hypothetical subject over 365 days of follow-up.

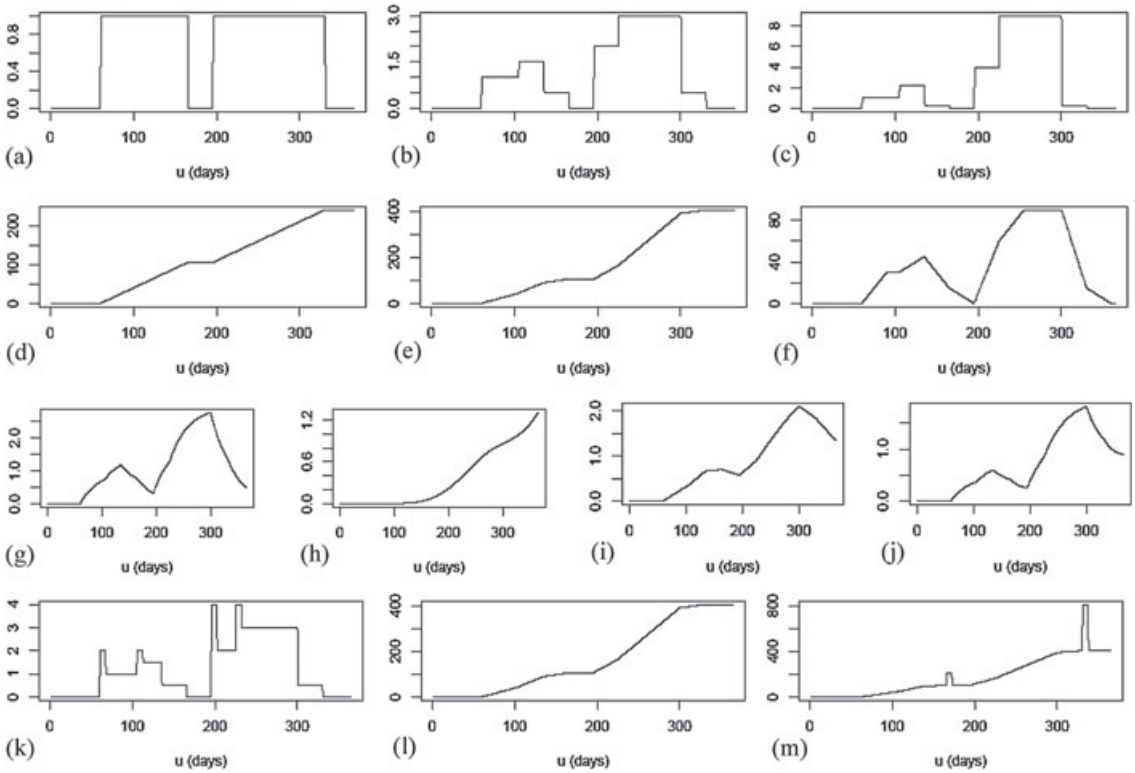

Figure 2. Exposure metric functions $M(u)$ corresponding to the daily dose pattern shown in Figure 1 for: (a) model 1 (current use); (b) model 2 (current dose); (c) model 3 (NL current dose); (d) model 4 (duration past use); (e) model 5 (cumulative past dose); (f) model 6 (cumulative dose past m days) where $m = 30$ days; (g) model 7 (WCE model with exponential weight function); (h) model 8 (WCE model with inverted U weight function); (i) model 9 (WCE model with truncated normal weight function); (j) model 10 (WCE model with weight function as the mixture of exponential and inverted U functions shown in panels (g) and (h)); (k) model 11 (increment and current dose); (l) model 12 (increment and cumulative past dose); (m) model 13 (withdrawal).

the past ($X(t)$, $t \leqslant u$) on current risk at time u. For example, model 2 (Figure 2(b)) assumes that only the current dose $X(u)$ affects the current risk, whereas the total cumulative dose model 5 (Figure 2(e)) assumes that all past doses have an equal impact on the current risk, regardless of their timing.

2.2. *Recency-weighted cumulative exposure model*

All conventional models 1 to 6 may be cast as special cases of a more general, recency-weighted cumulative exposure (WCE) model. In the WCE model, the etiologically relevant exposure metric is defined as a weighted function of past doses $\mathbf{X}(t), t = 1, \ldots, u$ [13–15]:

$$M(u) = \text{WCE}(u) = \sum [w(u-t)X(t), \text{for } t \leqslant u] \tag{2}$$

where $w(u-t)$ is the weight function quantifying the relative importance of past doses on current risk as a function of time elapsed between t and u.

The WCE metrics represent a subclass of $M(u)$ functions limited to additive linear functions of past doses. In other words, the time-dependent exposure metric WCE(u) is a weighted sum of past doses, $X(t)$ for $t \leqslant u$, each of which is multiplied by a weight $w(u-t)$, which may be either fixed a priori [15] or estimated from the data [13]. In principle, these weights could take any arbitrary values. However, it is plausible to assume that weights, which represent the relative importance assigned to consecutive doses, change with increasing time since exposure $u-t$ according to a smooth weight function $w(u-t)$.

Because the true weight function is typically unknown, we have recently proposed a flexible WCE model in which a smooth weight function is estimated using cubic B-splines with fixed knots [16]. The spline-based WCE model for time-to-event analyses can be implemented using artificial time-dependent covariates in the Cox PH model [16].

Depending on the biological mechanism that links past drug exposure with the current risk of the clinical endpoint of interest, the shape and the values of the weight function $w(u-t)$ may vary considerably. For example, if more recent doses have higher impact, then $w(u-t)$ will be a decreasing function of time since exposure $u-t$, with $w(\Delta t) = 0$ for $\Delta t > a$, indicating that drug doses taken more than a days ago have no impact and can be ignored when assessing current risk at time u. In some applications, $w(u-t)$ may be a priori constrained to be nonnegative for all values of $u-t$, which would imply that the effects of past exposure/doses on the current risk are all in the same direction (e.g., never protective), regardless of the time since the exposure. However, for some drugs, the effects of past exposure may change over time from protective (negative weights) to harmful (positive weights) because of such phenomena as habituation or withdrawal effects [17, 18] or different pathophysiological pathways underlying short-term versus long-term effects of the drug. In such cases, weights assigned to different exposure times should be allowed to take either positive or negative values. Furthermore, the estimated weights are *not* constrained to sum up to any prespecified value. In particular, the relative values of weights assigned to doses taken at different times in the past determine the shape of the estimated weight function, whereas the sum of these weights determines the estimated total impact of the cumulative past dose on the hazard [16].

Figure 3(a)–(c) shows weight functions implied by conventional models 1 to 6. Although weight functions for all these conventional models are step functions of time since exposure $u-t$, it seems more plausible that the impact of past doses changes as a *smooth* function of time. Indeed, the WCE model with a smooth weight function may fit real-life data substantially better than the conventional models [15].

Panels (d)–(g) of Figure 3 show four alternative smooth weight functions (exponential, inverted U and truncated normal distributions, and mixture of exponential and inverted U distributions) considered in simulations described in Section 3. The corresponding WCE models are referred as models 7 to 10. Panels (g)–(j) of Figure 2 show the corresponding $M(u)$ functions, for the exposure pattern shown in Figure 1.

2.3. *New two-parameter models accounting for recent dose changes*

In some applications, the current risk may be additionally affected by a recent increase of the dose [17, 18]. Accordingly, models 11 and 12 define the exposure measure $M(u)$ as two-effect function accounting for both recent dose increase and either current dose (model 11) or total (unweighted) cumulative past dose (model 12):

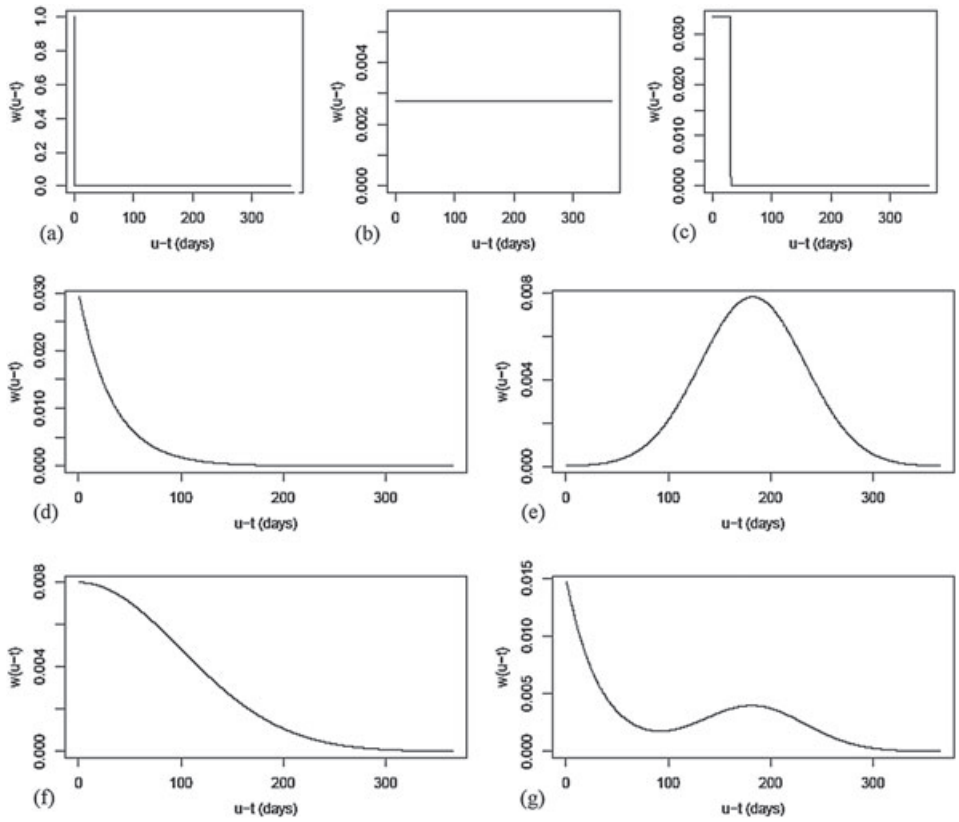

Figure 3. Weight functions $w(u-t)$ implied by models: (a) models 1–3: 1 (current use), 2 (current dose), and 3 (NL current dose); (b) models 4 and 5: 4 (duration past use) and 5 (cumulative past dose); (c) model 6 (cumulative dose past m), where $m = 30$ days; (d) 7 (WCE model with exponential weight function); (e) 8 (WCE model with inverted U weight function); (f) 9 (WCE model with truncated normal weight function); (g) 10 (WCE model with weight function as the mixture of exponential and inverted U function shown in panels (f) and (g)). All weight functions were normalized to have the total area under the curve equal to 1.0.

Model 11. Recent dose increment + current dose (increment and current dose): $M(u) = d(u) + X(u)$, where $d(u)$ is an indicator of an increase in dose in the last 7 days: $d(u) = [X(u) - X(u-7)]_+$, with $[z]_+ = 0$ if $z \leqslant 0$ or $[z]_+ = z$ if $z > 0$;
Model 12. Recent dose increment + total cumulative past dose (increment and cumulative past dose): $M(u) = d(u) + \sum[X(t), \text{for } 0 < t \leqslant u]$, where $d(u)$ is defined as for model 11.

Finally, the risks of some adverse events may be exacerbated by a recent withdrawal of the drug [17]. Moreover, possibly because of habituation effects [18], the impact of such withdrawal may increase with increasing cumulative past dose. Accordingly, model 13 defines the exposure measure $M(u)$ as a two-effect function depending on total (unweighted) cumulative past dose and its interaction with recent withdrawal:

Model 13. Recent withdrawal effect interacting with cumulative past dose (withdrawal): $M(u) = I[C(u)] \cdot \sum[X(t), \text{for } 0 < t \leqslant u] + \sum[X(t), \text{for } 0 < t \leqslant u]$, where $I[C(u)] = 1\, if\, X(u) = 0$ and $X(k) \neq 0$ for at least one $k \in [u-7; u-1]$ and is a binary indicator of the cessation of drug use within the last 7 days.

Panels (k)–(m) of Figure 2 show the $M(u)$ functions obtained by applying respectively models 11 to 13 to the exposure pattern shown in Figure 1.

3. Simulation studies

3.1. Simulation design, assumptions, and data generation

Study design. We simulated hypothetical pharmacoepidemiological studies of an adverse effect of a single medication, with 600 new drug users followed-up for 1 year since the first prescription (time 0). We censored subjects without event at 365 days and assumed no losses to follow-up. We generated event

times from the marginal exponential distribution, selected to obtain either 200 or 500 uncensored events in the first year of follow-up (66.7% or 16.7% censoring).

Exposure matrix. We generated time-varying patterns of drug use and dose ($\mathbf{X}_i(t), t = 1, \ldots,$ 365 days) for subjects $i = 1, \ldots, 600$. Subjects could repeatedly stop and, after a nonuse period ($X_i(t) = 0$), restart the drug use ($X_i(t) > 0$). The length of consecutive use and nonuse periods varied both between and within subject (Supplementary Materials, section 1.1).

Alternative exposure–risk models. To link exposure vectors ($\mathbf{X}_i(t), t = 1, \ldots, 365$ days) with event times, we considered 13 different models for $M(u)$, described in Section 2 (see also Figure 2), with $g(X(u)) = [X(u)]^2$ for the nonlinear current dose model 3.

To ensure the comparability of the results, we standardized the strength of the association across the 13 'true' models (Supplementary Materials, section 1.2).

We then assumed the following univariate Cox PH model:

$$\lambda[u|(\mathbf{X}_i(t), t = 1, \ldots, u)] = \lambda_0(u)exp[\beta^* \cdot M(u)] \tag{3}$$

where $M(u)$ is the value of the model-specific exposure metric, $\lambda_0(u)$ is the baseline hazard, corresponding to $M(u) = 0$, and β^* is the standardized logarithm of hazard ratio (HR) for 1 SD increase in $M(u)$.

Finally, for each of the 13 true models and each simulated sample, we matched individual simulated event or censoring times with individual dose time vectors ($\mathbf{X}_i(t), t = 1, \ldots, u$) using our permutational algorithm [19, 20], recently adapted and validated for generating event times conditional on complex time-varying exposures [21, 22] (Supplementary Materials, section 1.3).

Simulation scenarios. We simulated three main scenarios corresponding to the following: (i) 200 events and HR = 2.0; (ii) 500 events and HR = 2.0; and (iii) 500 events and HR = 1.5. For both sample sizes, we also simulated data assuming no association (HR = 1.0). For each simulated scenario, we generated 500 independent and identically distributed random samples for each of the 13 true models.

3.2. Analyses of simulated data

We analyzed the simulated data using a two-stage procedure. Stage 1 corresponded to 'exploratory analyses', aimed at identifying the best fitting among the alternative exposure–risk models considered. Then, stage 2 'confirmatory analyses' relied on the model selected at stage 1 to (i) test statistical significance of the association between the corresponding exposure metric and the hazard and (ii) estimate the exposure effects.

Estimation of alternative exposure–risk association models. We analyzed each simulated sample with 10 alternative versions of the univariate Cox model in Equation (3), which covered the range of true models used to generate the data. They included models 1 to 6 (Section 2.1) and models 11 to 13 (Section 2.3), as well as the flexible WCE model described in the next subsection. In model 3, we estimated the nonlinear effect of current dose model using cubic B-splines with 1 interior knot.

Estimation of flexible WCE models. Briefly, in each simulated sample, we estimated six alternative WCE models using our spline-based estimation method [16], corresponding to different combinations of (i) length a of time window over which past doses may affect current risk (a = 180 or 365 days), and (ii) 1–3 interior knots, placed at equal intervals within the $[0; a]$ time window [16]. We a priori constrained both the estimated weight functions and their first derivatives to equal 0 at $u - a$, that is $w(u-a) = 0$ and $w'(u-a) = 0$ [16]. The constrained cubic spline WCE models with 1–3 interior knots used, respectively, 3–5 degrees of freedom (DOF) to estimate the weight function in Equation (2). We selected the best-fitting 'final' WCE model for a given simulated sample using the minimum AIC or BIC criterion. Section 1.4 of the Supplementary Materials provides more details on WCE model estimation.

Model discrimination. To compare the fit of the 10 alternative models estimated for a given simulated sample, we employed both the AIC [23] and the BIC adapted to survival analysis [24] criteria:

$$\text{AIC} = D + 2k \tag{4a}$$

$$\text{BIC} = D + \log(m) \cdot k \tag{4b}$$

where D is the deviance of the respective model, calculated by multiplying the corresponding maximized partial log likelihood by -2, k is the number of model's DOF, and m is the number of uncensored events. Notice that $k = 1$ for simple conventional parametric models 1 and 2 and 4–6, $k = 2$ for the two-parameter models (models 11–13 in Section 2.3) and $k = 4$ for the cubic spline estimate of the

nonlinear effect of current dose (model 3). Finally, when calculating the DOF of the constrained WCE model, selected for a given sample, we additionally penalized for a posteriori data-dependent choice of the best fitting among the six alternative models. Accordingly, we set $k = 4$, 5, or 6 if the best-fitting WCE model used 1, 2, or 3 interior knots.

We then selected the minimum AIC and BIC for a given sample and evaluated their performance in identifying the 'true' model. To this end, we estimated the proportion of samples where the true model, from which the data for a given scenario were generated, corresponded to the best-fitting model, or diverged from the optimal model by less than 4 AIC or BIC points (such minor differences in AIC or BIC will be likely interpreted as equally good fit of the corresponding models [25]).

Testing the association conditional on the best-fitting model. For each simulated sample, we tested the null hypothesis of no association between the corresponding model-specific exposure metric $M(u)$ and the hazard, separately for each of the 10 estimation models, at the significance level of $\alpha = 0.05$. Assessment of the performance of the test, using a specific model, across *all* samples simulated for a given scenario, helps in investigating potential loss of power due to a priori selection of an 'incorrect' model, that is, *not* consistent with the 'true' model used to generate the data.

However, in our two-stage analyses, the tests performed at the second (confirmatory) stage are *conditional* on the best-fitting model, selected a posteriori through stage 1 exploratory analyses. To avoid type I error inflation [26], we corrected the critical values of such conditional tests using a computationally intensive hypothesis testing procedure [16] (section 1.5 of the Supplementary Materials provides details).

We estimated the empirical power of all *unconditional* tests as the proportion of samples in which the corresponding likelihood ratio test (LRT) yielded $p < 0.05$. In simulations where the exposure had *no* association with the hazard, we estimated the empirical type I error rates as the proportion of 5000 samples where the p-value of the model-specific test was below the cut-off of 0.05 for unconditional tests and below the corrected simulation-based p-value for conditional tests.

We implemented all steps of data generation and analyses using R [27].

4. Simulation results

4.1. Model discrimination

Table I shows the proportion of simulated samples where the alternative goodness-of-fit criteria identified the true data-generating model (columns 1 and 2) as the best-fitting model, separately for the three combinations of the number of events and the association strength (HR). The percentages in brackets indicate how often the true model yielded AIC or BIC within 4 points of the best-fitting model, which may be interpreted as an approximately equivalent fit [25]. If the true model was one of the WCE models 7 to 10, it was deemed to be correctly identified if the best-fitting model corresponded to one of the cubic spline-based WCE estimators. As expected, all criteria are more successful in identifying the true models when the exposure effect becomes stronger (HR = 2.0 vs. HR = 1.5) and the number of uncensored events increases (500 vs. 200).

Table I indicates also that the relative advantages of AIC versus BIC criteria depend on the complexity of the true model. BIC more frequently identified simpler conventional 1 DOF parametric models, whereas AIC was more much likely to correctly identify complex WCE models 7 to 10, with 4 to 6 DOF, and the nonlinear dose model 3. This contrast occurred because the BIC-induced penalty for one additional DOF is log(number of events) [24], that is, 5.30 and 6.21 for respectively 200 and 500 events, whereas the AIC-induced penalty equals only 2, regardless of the number of events [23].

Overall, the proportion of correctly identified true models was somewhat higher for AIC than for BIC. For example, with 500 events and (stronger) HR = 2.0, each of the 13 true models was selected by AIC in at least 64.0% of samples, whereas BIC identified some of the WCE models in only 18.0% or 28.4% of samples. Furthermore, even with only 200 events, each true model was within 4 points of the best AIC model in more than 88% of samples. In contrast, the more complex true models (nonlinear current dose model 3 or WCE models 7–10) were within 4 points of the minimum BIC model in less than 55% of samples for the same scenario (Table I). See section 1.6 of the Supplementary Materials for further discussion.

4.2. Type I error rates and accuracy of effect estimates

In simulations where we assumed drug exposure had no impact on the hazard, most unconditional tests yielded—as expected—type I error rates close to the nominal $\alpha = 0.05$ (data not shown). In contrast, the

Statistics in Medicine

Table I. Percentage of samples with true data-generating model identified as the best-fitting model by AIC or BIC criteria.

		Scenario 1: 200 events and HR = 2.0		Scenario 2: 500 events and HR = 2.0		Scenario 3: 500 events and HR = 1.5	
Model #	True model	Best AIC (≤4 from best AIC)	Best BIC (≤4 from best BIC)	Best AIC (≤4 from best AIC)	Best BIC (≤4 from best BIC)	Best AIC (≤4 from best AIC)	Best BIC (≤4 from best BIC)
1	Current use	89.6 (97.6)	99.6 (99.8)	89.2 (99.0)	100 (100)	86.2 (97.6)	100 (100)
2	Current dose	74.6 (96.2)	97.4 (99.6)	79.0 (97.8)	99.2 (100)	76.2 (97.2)	98.6 (99.6)
3	NL current dose	74.8 (96.4)	18.8 (32.6)	97.8 (100)	64.0 (80.2)	69.6 (94.6)	8.2 (17.6)
4	Duration past use	85.8 (97.2)	89.0 (98.6)	95.2 (97.8)	96.4 (98.2)	77.0 (95.4)	81.6 (97.0)
5	Cumulative past dose	60.0 (94.8)	92.2 (98.8)	64.0 (94.2)	95.8 (99.6)	58.2 (92.8)	87.4 (98.0)
6	Cumulative dose past 30 days	85.6 (95.6)	98.0 (99.6)	89.0 (98.2)	100 (100)	86.2 (97.0)	99.0 (99.4)
7	WCE exponential	57.6 (88.6)	5.8 (17.0)	90.4 (98.4)	18.0 (39.0)	52.4 (89.4)	1.0 (5.2)
8	WCE inverted U	79.6 (95.6)	24.8 (43.0)	96.4 (99.8)	50.4 (71.2)	59.8 (86.8)	4.2 (10.6)
9	WCE truncated N	84.2 (98.2)	31.0 (53.2)	97.8 (99.6)	55.8 (73.6)	63.0 (91.0)	2.8 (11.8)
10	WCE mixture	65.6 (93.0)	5.4 (19.0)	96.0 (99.6)	28.4 (56.2)	40.0 (85.0)	0.2 (1.4)
11	Increment and current dose	98.6 (100)	95.2 (99.2)	100 (100)	100 (100)	96.2 (99.8)	76.0 (96.2)
12	Increment and cumulative past dose	93.4 (99.4)	84.4 (96.8)	99.8 (100)	100 (100)	88.6 (99.0)	67.8 (89.9)
13	Withdrawal	99.8 (100)	99.8 (100)	100 (100)	99.8 (100)	81.6 (99.0)	59.0 (82.2)

HR, hazard ratio; WCE, weighted cumulative exposure.

conditional tests, based on the model minimizing AIC in the corresponding sample, had highly inflated type I error rates of 0.22 to 0.28. The proposed computer intensive testing procedure [16], based on 5000 random samples (Supplementary Materials, section 1.5), yielded the corrected significance levels of 0.0084 and 0.0083 for simulations with respectively 200 and 500 events per sample. The closeness of the two corrected significance levels suggests that the correction for conditional testing is robust to the effective sample size. When we applied the corrected significance levels to 5000 new samples, generated assuming no association, the resulting empirical type I error rates were 0.0524 and 0.0516, that is, very close to the 0.05 rate.

A posteriori model selection may also result in biased conditional effect estimates obtained at the confirmatory stage of the analyses [28]. In our simulations, such bias was reflected in the finding that 95%CI for the bias of the conditional estimates for simulated samples where the respective true model was selected by either AIC or BIC often excluded 0 for the one-parameter true models. However, for all scenarios, the relative bias was below 8%, and it did not exceed 5% in almost all situations (data not shown). Similarly, in all simulations where the true model was one of the four WCE models 7–10, the mean conditional weight function estimates, based on only those samples where the WCE model was selected by either AIC or BIC criteria, reflected well the shape of the true weight function (data not shown). These findings indicate that, in our specific context, a posteriori model selection had only very minor impact on the accuracy of the final effect estimates. This reflects the ability of AIC and BIC to correctly identify the true model in a majority of simulated samples (Table I), which minimizes the impact of selection bias on the results of confirmatory analyses.

4.3. Impact of assuming an incorrect model on power to detect a significant association

Table II compares, for each true model (rows), the empirical power of unconditional tests, which are based on selected estimation models, to reject the null hypothesis of no association. Specifically, for each of the three combinations of the number of events and exposure effect and each true model, we report the power of the tests for the fitted models yielding to the lowest and the second lowest power and for the final (minimum AIC) flexible WCE model. The power is reported for either the Wald test (for 1 DOF estimators) or LRT (for the nonlinear current dose model 3 with 4 DOF, the WCE estimator with 4, 5, or 6 DOF, and the two-parameter models 11–13 with 2 DOF), using two-tailed $\alpha = 0.05$ for both tests.

In almost all situations, the power for the unconditional test for the respective true model exceeded 89.5%. Suboptimal power (41.2%–86.8%) occurred only when the true model was the inverted U WCE

Table II. Comparison of power for the tests of no association for the different models fitted*.

		Scenario 1: 200 events and HR = 2.0		
Model no.	True model	Lowest power (%)	Second lowest power (%)	Power best WCE model** (%)
1	Current use	Cumulative past dose (18.6)	Increment and cumulative past dose (23.0)	45.6
2	Current dose	Duration past use (54.8)	Cumulative past dose (84.4)	100
3	NL current dose	Duration past use (47.2)	Withdrawal (90.6)	100
4	Duration past use	NL current dose (6.8)	Increment and current dose (12.2)	20.8
5	Cumulative past dose	Current use (20.0)	NL current dose (20.4)	92.8
6	Cumulative dose past 30 days	Duration past use (77.8)	Withdrawal (96.2)	100
7	WCE exponential	Duration past use (90.0)	Current use (94.4)	100
8	WCE inverted U	NL current dose (3.0)	Current use (3.4)	62.8
9	WCE truncated N	Current use (56.2)	NL current dose (76.2)	100
10	WCE mixture	Current use (69.8)	NL current dose (90.8)	99.4
11	Increment and current dose	Duration past use (31.8)	Cumulative past dose (58.4)	97.2
12	Increment and cumulative past dose	Current use (30.8)	NL current dose (44.2)	84.4
13	Withdrawal	Current dose (8.6)	Increment and current dose (12.6)	94.4

Table II. *Continued.*

Model no.	True model	Scenario 2: 500 events and HR = 2.0 Lowest power (%)	Second lowest power (%)	Power best WCE model** (%)
1	Current use	Cumulative past dose (71)	Increment and cumulative past dose (81)	99.2
2	Current dose	Duration past use (97.6)	100 for all other models	100
3	NL current dose	Duration past use (93.6)	100 for all other models	100
4	Duration past use	NL current dose (22.2)	Increment and current dose (34.4)	72.8
5	Cumulative past dose	Current use (42.6)	NL current dose (49.2)	99.8
6	Cumulative dose past 30 days	Duration past use (99.8)	100 for all other models	100
7	WCE exponential	100 for all models	100 for all models	100
8	WCE inverted U	NL current dose (2.8)	Current use (4.0)	86.8
9	WCE truncated N	Current use (91.2)	NL current dose (99.8)	100
10	WCE mixture	Current use (98.4)	100 for all other models	100
11	Increment and current dose	Duration past use (80.4)	Cumulative past dose (97.2)	100
12	Increment and cumulative past dose	Current use (67.8)	Cumulative dose past 30 days (83.2)	99.6
13	Withdrawal	Current use (5.4)	NL current dose (17.2)	100

Table II. *Continued.*

Model no.	True model	Scenario 3: 500 events and HR = 1.5 Lowest power (%)	Second lowest power (%)	Power best WCE model** (%)
1	Current use	Cumulative past dose (44.0)	Increment and cumulative past dose (47.2)	80.2
2	Current dose	Duration past use (71.4)	Withdrawal (86.8)	100
3	NL current dose	Duration past use (51.4)	Withdrawal (84.2)	100
4	Duration past use	NL current dose (8.6)	Incr. & current dose (14.2)	21.6
5	Cumulative past dose	NL current dose (15.6)	Current use (21.4)	78.0
6	Cumulative dose past 30 days	Duration past use (86.2)	Increment and cumulative past dose (97.4)	100
7	WCE exponential	Duration past use (93.8)	Current use (95.8)	100
8	WCE inverted U	NL current dose (3.6)	Current use and Current dose (6.0)	41.2
9	WCE truncated N	Current use (51.6)	NL current dose (64.0)	99.6
10	WCE mixture	Current use (62.6)	NL current dose (79.4)	97.4
11	Increment and current dose	Duration past use (51.8)	Withdrawal (70.8)	98.6
12	Increment and cumulative past dose	Current use (28.4)	NL current dose (28.8)	67.0
13	Withdrawal	Current use (4.6)	NL current dose (7.2)	72.8

HR, hazard ratio; WCE, weighted cumulative exposure; NL, nonlinear.

*Empirical power of the test of no association, using $\alpha = 0.05$, estimated from 500 simulated samples. In almost all situations, the power based on the respective true model was greater than 97.5%. The only exceptions concerned the inverted U WCE model with model 8 (powers of 62.8%, 86.8%, and 41.2% respectively for scenarios 1, 2, and 3) and the duration past use model 4 (power of 89.6% for scenario 3).

**Best WCE model identified by AIC. DOF of the LRT account for 1 DOF penalty for selection of the best time window and number of knots.

model (model 8), for which exposure starts affecting the risk only several months later (Figure 3(b)). On the other hand, the loss of power due to relying on the incorrect model varies substantially across models (Table II). For example, when we generated data from the current use model 1, the tests based

on models 4 and 5 that incorrectly assumed long-term cumulative effects induced a dramatic loss of power, especially in simulations with weaker exposure effect or smaller number of events. Conversely, if the true model corresponds to unweighted cumulative use/dose models 4–5, then relying on any of the current-effect models 1–3 induced a substantial loss of power. Interestingly, regardless of the true model, the unconditional tests based on the flexible WCE model had always either a power close to 100% or at least twice as high as the test with the weakest power (Table II).

4.4. Evaluation of cubic spline-based weighted cumulative exposure estimators

Section 2.2 explains that the WCE model incorporates several simpler conventional pharmacoepidemiological models as its special cases and thus should be considered a plausible candidate model in many real-life applications. Therefore, it is important to assess the performance of WCE estimates. Each panel of Figure 4 shows a random sample of 100 AIC-optimal weight function estimates corresponding to a specific 'true' weight function $w(u-t)$ (four columns), with either 200 (upper row) or 500 events (lower row).

As expected, some flexible weight function estimates show considerable wiggliness, which reflects the overfit bias due to the instability of spline estimates, especially in the regions with little data [29]. Indeed, the weight estimates are often unstable near time 0, especially when the true effect of current dose is weak or nil (Figure 4(c) and (g) or (b) and (f), respectively). Because of this instability, the estimated weight functions $w(u-t)$ often take negative values over some interval of time since exposure $u-t$ (Figure 4), even if the true weight functions used to generate the data are always nonnegative (Figure 3(d)–(g)). However, because the negative weight function estimates in Figure 4 occur over only a limited time interval and the corresponding 95% pointwise confidence bands almost always include 0 (data not shown), they can be interpreted as not different from 0.

Furthermore, for the complex bimodal weight function, some estimates 'miss' the later, low 'peak' (Figure 4(d) and (h)), and incorrectly suggest that the doses taken more than 180 days ago have no impact on current risk. Similarly, Figure 4(b) and (f) shows underestimation of the impact of exposures that occurred about 200 days ago. Both findings reflect limited information available to estimate long-term exposure effects in a study with only 1-year follow-up. However, for all four weight functions and both event numbers, most estimates do reflect the general shape of the corresponding true weight function (Figure 4), providing a reasonably accurate assessment of the impact of past doses on the current hazard.

Table III shows the frequency with which the WCE model either yielded the minimum AIC or (in brackets) was within 4 points of the minimum AIC model. This indicates that the flexible WCE

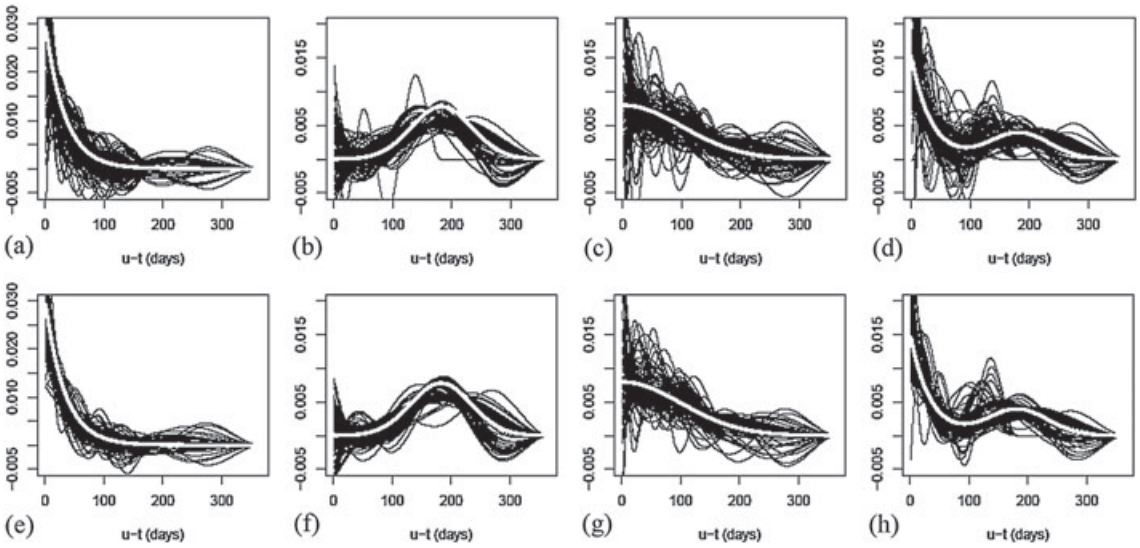

Figure 4. One hundred AIC-optimal randomly selected weight function estimates $\hat{w}(u-t)$ from fitted WCE models, for each true weight function considered (shown in white) for true WCE models. Panels (a)–(d) are for scenario 1 with 200 events and HR = 2.0, and panels (e)–(h) are for scenario 2 with 500 events and HR = 2.0. For comparison purposes, all curves were normalized to have an area under the curve equal to 1.0 (when considering the absolute values of $\hat{w}(u-t)$).

Table III. Percentage of samples for which the flexible weighted cumulative exposure model was identified as the best AIC or was within 4 points of the minimum AIC model (in brackets).

Model no.	True model	Scenario 1: 200 events and HR = 2.0	Scenario 2: 500 events and HR = 2.0	Scenario 3: 500 events and HR = 1.5
1	Current use	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
2	Current dose	3.2 (6.0)	0.2 (0.4)	1.4 (5.6)
3	NL current dose	1.0 (2.0)	0.0 (0.0)	0.8 (3.4)
4	Duration past use	3.6 (9.2)	0.6 (4.2)	2.0 (15.4)
5	Cumulative past dose	21.0 (39.8)	6.8 (38.2)	6.6 (40.8)
6	Cumulative dose past 30 days	25.4 (46.2)	11.0 (38.2)	11.6 (48.4)
7	WCE exponential	75.0 (88.6)	90.4 (98.4)	52.4 (89.4)
8	WCE inverted U	88.2 (95.6)	96.4 (99.8)	59.8 (86.8)
9	WCE truncated N	93.4 (98.2)	97.8 (99.6)	63.0 (91.0)
10	WCE mixture	81.4 (93.0)	96.0 (99.6)	40.0 (85.0)
11	Increment and current dose	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
12	Increment and cumulative past dose	2.0 (6.0)	0.2 (0.2)	1.8 (10.2)
13	Withdrawal	0.2 (0.2)	0.0 (0.0)	1.6 (13.0)

HR, hazard ratio; NL, nonlinear; WCE, weighted cumulative exposure.

model may often 'compete' with unweighted cumulative dose models 5 and 6 and less frequently with cumulative duration model 4 or current dose model 2 but is very unlikely to fit well the data generated from the other models (Table III).

Figure 5 is similar to Figure 4, except it shows the AIC-optimal WCE estimates for four selected cases where the true model does *not* correspond to any smooth WCE model. For data generated from the total (unweighted) past dose model 5, most flexible WCE estimates correctly suggest that the weights remain approximately constant across the entire follow-up period (Figure 5(b) and (f)). In the three other cases, the true exposure effect lasts only 1 to 30 days and the corresponding graphs are limited to 90 days. In Figure 5(a) and (e), weights estimated for data generated from the current dose model 2 exhibit an immediate sharp decline in the first few days, which correctly suggests an acute, very short-term effect. In contrast, when the true cumulative effect lasts for 30 days (model 6), most of spline estimates decrease at a much slower rate and correctly suggest that past doses taken up to about 1 month ago have a marked

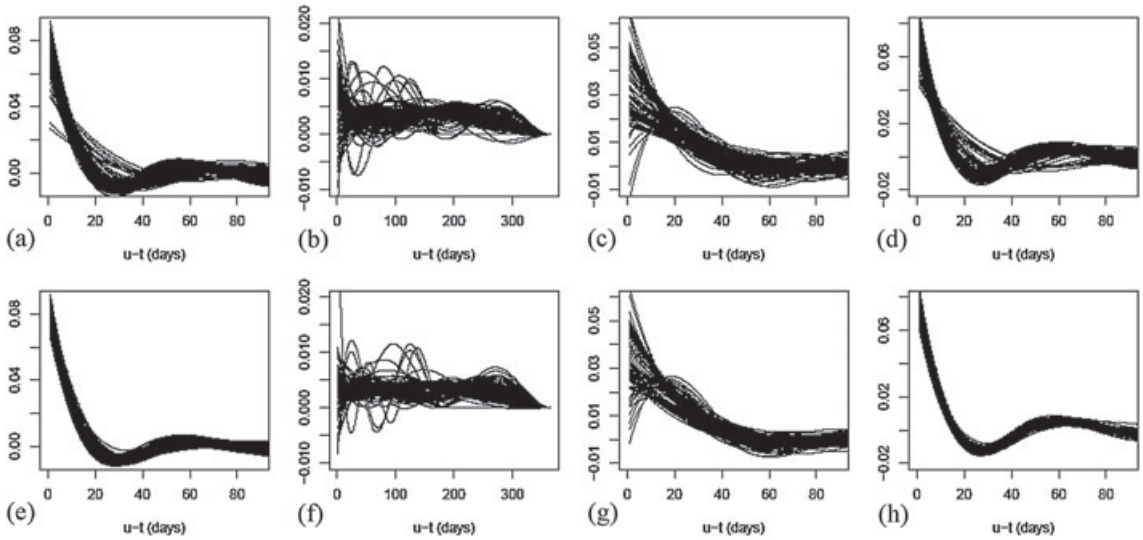

Figure 5. One hundred AIC-optimal randomly selected weight function estimates $\hat{w}(u-t)$) from fitted weighted cumulative exposure models for: (a) and (e) true model 2 (current dose); (b) and (f) true model 5 (cumulative past dose); (c) and (g) true model 6 (cumulative dose past 30 days); (d) and (h) true model 11 (increment and current dose). Panels (a)–(d) are for scenario 1 with 200 events and HR = 2.0, and panels (e)–(h) are for scenario 2 with 500 events and HR = 2.0. For comparison purposes, all curves were normalized to have an area under the curve equal to 1.0 (when considering the absolute values of $\hat{w}(u-t)$).

impact, whereas those taken longer ago have much weaker effects (Figure 5(c) and (g)). Even if, similar to Figure 4, some estimates show instability in the tails and/or implausible negative point weights, overall these results suggest that flexible WCE estimates approximated reasonably well the etiologically relevant exposure time window even if the true model corresponded to unsmooth (step function) weight functions assumed by conventional models, such as models 2, 5, and 6, which represent special cases of the general WCE model (Section 2.2).

In contrast to models 1–6, the two-parameter model 11, with separate effects of current dose and recent dose increase, is *not* a special case of the WCE model. Interestingly, the corresponding flexible weight function estimates not only (i) decrease sharply but even (ii) tend to assign *negative* weights to doses taken 10–40 days ago (Figure 5(d) and (h)). Feature (i) reflects the acute effect of the current dose (see previous discussion), whereas feature (ii) may be attributed to the fact that (a) subjects who had higher doses in the past 2–5 weeks were unlikely to have further increased their doses in the last week. On the other hand, in model 11, (b) recent dose increase is associated with an important risk increase. Conjunction of (a) and (b) implies that, among subjects with the same current dose $X(u)$, those who were exposed 2–5 weeks ago (implying $X(t) > 0$ for $u - 7 < t < u - 35$) have, on average, a *lower* risk than subjects who just started exposure ($X(u) > 0$ with $X(t) = 0$ for some $t < u - 7$), that is, who *increased* their dose in the last week. This lower risk associated with nonzero doses taken 2–5 weeks ago would explain why the corresponding weights in Figure 5(d) and (h) are *negative.* Obviously, the WCE estimates cannot identify the true model generating data as the two-parameter model 11, but the negative estimated weights may stimulate searching for clinically plausible explanations. Thus, even when the true model is *not* a special case of the WCE model, our flexible estimates may sometimes help with model diagnostics.

5. Real-life example

5.1. Methods

We reassessed the association between use of a psychotropic drug (flurazepam) and risk of fall-related injury among the elderly [30]. The original cohort included community-dwelling residents of the province of Québec, Canada, older than 65 years [2]. The present analyses were limited to 5111 subjects who started using flurazepam in 1990–1994 and were followed-up until December 31, 1994.

We estimated 10 alternative Cox PH models with flurazepam exposure history modeled by appropriate time-dependent covariates, corresponding to the 10 models fitted in simulations (Section 3.2). Given that the half-life elimination time of flurazepam from the plasma is between 40 to 100 h. [31], we a priori fixed the time window to 30 days for WCE modeling, while using 1 to 3 knots, as in our simulations. Event time corresponded to time from the first flurazepam prescription to the first fall-related injury [2, 16]. We censored individuals who died, moved out of the province, or were institutionalized before their first fall-related injury. We also censored subjects who had the event during hospitalizations when drug exposure was unknown. For other subjects, we set daily dose during hospitalization(s) to 0.

5.2. Results

Among the 5111 flurazepam users, followed-up for up to 5 years, 264 (5.2%) had fall-related injuries. Table IV compares the AIC for the 10 alternative models and reports the model-specific p-values for testing the null hypothesis of no association between respective measures of flurazepam exposure and fall-related injuries. The spline-based WCE model with 3 knots and the withdrawal model 13 (Section 2.3) yielded the best and second-best AIC values and indicated very significant risk increases. In contrast, almost all other models, with considerably higher AIC values, yielded nonsignificant results (Table IV).

Table IV. Comparison of AIC and tests of association for alternative models linking flurazepam use with risk of fall-related injuries.

Model no.	1	2	3	4	5	6
True model	Current use	Current dose	NL current dose	Duration past use	Cumulative past dose	Cumulative dose past 30 days
AIC	4118.2	4116.9	4118.0	4118.3	4118.2	4113.2
p-value	0.8225	0.2263	0.3730	0.9452	0.7671	0.0164

Table IV. *Continued.*				
Model no.	7–10	11	12	13
True model	WCE with 3 knots & 30-day window*	Increment and current dose	Increment and cumulative past dose	Withdrawal
AIC	4102.7	4118.9	4120.1	4106.3
p-value	0.0001	0.5008	0.8959	0.0009

NL, nonlinear; WCE, weighted cumulative exposure.
*Best WCE model identified by AIC. DOF of the LRT account for 1 DOF penalty for the selection of number of knots.

As expected, based on short half-life, doses taken more than 10 days ago have no impact on the current risk. Interestingly, the weights functions for all WCE model assigned *negative* weights to flurazepam doses from the last 2–3 days (Figure 6). In contrast, doses taken 4–7 days ago were assigned high positive weights.

To assess the robustness of the negative weight estimates, we a posteriori constrained the best-fitting 3-knot WCE model to have 0 values for the estimated weight function and its first derivative at time 0 (see [16] for details). This additional (nonnegativity) constraint resulted in a statistically significant loss of fit to data (2 DOF LRT statistics with $p < 0.0001$). This confirmed that most recent flurazepam doses should be, indeed, assigned negative weights. The negative weights for most recent doses, together with positive weights for previous 4–7 days, imply that the risk is the highest for subjects who recently stopped taking flurazepam and had high doses in the previous week. This may suggest a *withdrawal effect* [17]. Indeed, the second best-fitting withdrawal model 13 (Table IV) indicated a very significant interaction between recent withdrawal and past cumulative flurazepam dose ($p < 0.0001$). The interaction coefficient (HR = 1.0064, 95%CI: 1.0040, 1.0088) was much stronger than the (nonsignificant) cumulative dose effect in the absence of withdrawal, indicating that withdrawal in the past week dramatically increases the impact of past cumulative dose. These results are consistent with the negative weights assigned to the last few days by the WCE models. Although further studies are necessary to confirm if the increased risk of fall-related injuries among recent flurazepam users may be partly due to withdrawal effects, this conjecture would also explain why the current use and current dose models 1–3 yielded high AIC and nonsignificant results (Table IV). Overall, this real-life example illustrates the potential benefits of considering alternative exposure models, including the flexible WCE model [16], to both detect an association between drug exposure and risk of an adverse event and get new insights regarding the underlying biological mechanisms.

The WCE model implies that different vectors of past doses $\mathbf{X}(t), t \leqslant u$, correspond to different current hazards (at time u). See [16] for formulas that show how the WCE estimates can be transformed into estimates of HRs for the different vectors of past doses. To illustrate the clinical implications of the weight function estimated with the best-fitting WCE model (bold curve in Figure 6), Table V shows the adjusted HRs for different patterns of recent flurazepam exposure, relative to a 'nonuser' who did not take the drug in the last 30 days. The last row shows that doses taken more than 1 week ago had no effect on the current risk, which also explains why the HR does not increase if current exposure started 2 weeks rather than 1 week ago (two first rows of Table V). For more recent exposure patterns, doubling the daily dose from one-half (dose = 0.5) to the WHO recommended dose (dose = 1) only had

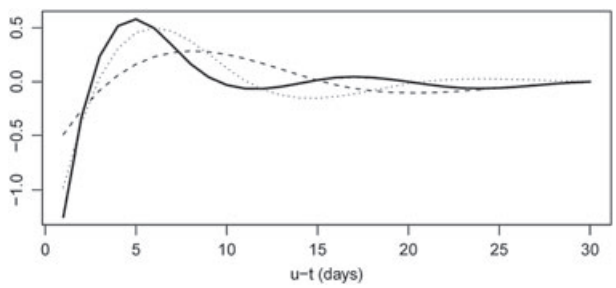

Figure 6. Weight function estimates $\hat{w}(u-t)$ for WCE models with 30-day window for flurazepam users: bold curve for best-fitting WCE (3 knots), dashed curve for WCE model with 1 knot, and dotted curve for WCE model with 2 knots.

Table V. Hazard ratios for specific patterns of exposure for the best-fitting weighted cumulative exposure model.

		Hazard ratio	
Pattern of exposure	Reference	Dose = 0.5	Dose = 1.0
Current user for 2 weeks	Nonuser	1.33	1.76
Current user for 1 week	Nonuser	1.34	1.79
Past user for 1 week, who stopped 3 days ago	Nonuser	2.86	8.21
Past user for 1 week, who stopped 1 week ago	Nonuser	0.99	0.98

a moderate impact on current users but produced a high point estimate of HR $>$ 8.0 for a past user who stopped the treatment 3 days ago and thus who may be affected by the short-term withdrawal effect (see previous discussion).

6. Discussion

Current pharmacoepidemiology studies often select a priori a given conventional model to represent the impact of drug exposure on the risk of a particular adverse event and do not attempt to assess its consistency with empirical data [5–8]. However, pharmacokinetic and pharmacodynamic characteristics of different drugs, as well as different physiological responses to different pharmacological agents, imply that drug exposure–risk associations may vary substantially across different medications and various adverse effects [9]. Although the 'true' model for a particular exposure–outcome association is rarely known, our simulations demonstrate that statistical goodness-of-fit criteria help identifying model(s) approximating reasonably well the underlying data-generating mechanism. Indeed, in our simulations, the minimum AIC criterion identified each of the true models in a majority of simulated samples, even with only 200 or 500 observed events, in almost all cases. Furthermore, in the vast majority of the remaining samples, the true model was within 4 AIC points of the best-fitting model and would be considered as a viable alternative [25]. Even if AIC may sometimes produce inconsistent results, in our simulations, it performed somewhat better than BIC (Table I). BIC induces excessive penalties for the number of events, especially in large studies, often leading to a choice of overly simple models (Section 4.1). Thus, we tentatively suggest using AIC criterion to identify the 'correct' model. Still, we recommend that investigators report both AIC and BIC of alternative models and consider both criteria while selecting one or few best-fitting models for their specific application. Further analytical work and simulations will help understanding under which conditions AIC, BIC, or their modifications [32] will be most accurate in identifying the correct exposure model among alternative models for $M(u)$ in Equation (1).

There are two main reasons why we recommend that pharmacoepidemiological studies should consider different statistical models and compare their fit to empirical data to identify the (approximately) correct model. A priori selection of an incorrect model may substantially reduce the power for testing the adverse drug effects (Table II). Even more importantly, as shown in Figure 2, different models may lead to very different risk prediction and, thus, to different guidelines regarding prescribing patterns expected to minimize the risk of adverse events.

In this context, we believe that the WCE model [16] should be routinely considered in pharmacoepidemiological studies. The WCE model incorporates many conventional pharmacoepidemiological models as its special cases. Furthermore, our flexible WCE model [16] has two important advantages: (i) it avoids arbitrary a priori assumptions regarding the time window of etiologically relevant exposure; and (ii) by estimating a *smooth* weight function, it yields more clinically plausible results than the conventional step-function-based models. Our cubic spline weight function estimates permitted a reasonably accurate assessment of the relative importance of doses taken at different times in the past, even if—especially with a limited number of events—they may exhibit some overfit bias (Figure 4), typical of very flexible estimates [33–35]. Further research is necessary to assess if, and for what data structures and sample sizes, choosing the final spline model on the basis of a more conservative criterion, such as BIC or its modifications [32, 36] rather than AIC, may improve the accuracy of the estimates. Interestingly, simulations also suggest that our flexible estimates may help with model diagnostics even if the true model is *not* a WCE model (Figure 5).

Some limitations of our study have to be recognized. Firstly, the 10 models fit in our analyses represent only a subset of models that may apply in pharmacoepidemiology [9]. Although we included some of the

most popular conventional models, it is difficult to even conceive the finite list of *all* potentially relevant models. Secondly, to reduce computational burden, we considered only a moderate number of events in the simulations. Modern pharmacoepidemiology relies mostly on large administrative databases [1], often with much larger numbers of events [37–39] but possibly with weaker associations. Furthermore, in any real-life application, confounding by indication [40–42] and/or errors in exposure measurement [43] should be considered. The way such data limitations may affect estimates from particular exposure models will likely vary across applications. For example, if many drug users interrupt the treatment after, say, T days, the estimated weight function may assign artificially low weights to doses assumed to be taken T or more days ago. To account for plausible patterns of noncompliance, future research should consider adapting exposure measurement models, recently extended to time-varying exposures in survival analyses [44], to WCE modeling. However, it will be difficult to rely on weight function estimates to infer about *unknown* actual patterns of noncompliance. For example, the same low weights assigned to doses taken more than T days ago may simply reflect the fact that, because of pharmacodynamic properties of a given drug, the actual time window over which past doses affect current risk does not exceed T days. On the other hand, it is rather unlikely that specific patterns of noncompliance will seriously affect the comparison of the fit with data of alternative exposure models. In summary, whereas future studies should consider a wider range of models, larger sample sizes, and the aforementioned data limitations, we believe our simulations illustrate well the methodological issues of primary interest for the current study.

Methodological issues addressed in our study also apply to many epidemiological studies of different time-varying exposures. Indeed, the WCE approach was initially proposed for case-control studies of occupational and environmental exposures [13, 14], and such applications motivated some recent, elegant developments in this area [45, 46]. We hope that our encouraging results will stimulate wider applications of the WCE modeling in (pharmaco-) epidemiological studies of time-varying exposures.

Acknowledgements

We are thankful to Dr. Robyn Tamblyn for providing us with the data set for the real-life example. Michal Abrahamowicz is a James McGill Professor at McGill University. A grant from the Canadian Institutes for Health Research (MOP-81275) supported this research.

References

1. Skegg DCG. Evaluating the safety of medicines, with particular reference to contraception. *Statistics in Medicine* 2001; **20**:3557–3569.
2. Tamblyn R, Abrahamowicz M, du Berger R, McLeod P, Bartlett G. A 5-year prospective assessment of the risk associated with individual benzodiazepines and doses in new elderly users. *Journal of the American Geriatrics Society* 2005; **53**:233–241.
3. Pilote L, Abrahamowicz M, Rodrigues E, Eisenberg MJ, Rahme E. Mortality rates in elderly patients who take different angiotensin-converting enzyme inhibitors after acute myocardial infarction: a class effect? *Annals of Internal Medicine* 2004; **141**:102–112.
4. Bartlett G, Abrahamowicz M, Tamblyn R, Grad R, Capek R, du Berger R. Longitudinal patterns of new benzodiazepine use in the elderly. *Pharmacoepidemiology and Drug Safety* 2004; **13**:669–682.
5. Smitten AL, Choi HK, Hochberg MC, Suissa S, Simon TA, Testa MA, Chan KA. The risk of hospitalized infection in patients with rheumatoid arthritis. *Journal of Rheumatology* 2008; **35**:387–393.
6. Franklin J, Lunt M, Bunn D, Symmons D, Silman A. Risk and predictors of infection leading to hospitalisation in a large primary-care-derived cohort of patients with inflammatory polyarthritis. *Annals of the Rheumatic Diseases* 2007; **66**:308–312.
7. Wolfe F, Caplan L, Michaud K. Treatment for rheumatoid arthritis and the risk of hospitalization for pneumonia: associations with prednisone, disease-modifying antirheumatic drugs, and anti-tumor necrosis factor therapy. *Arthritis and Rheumatism* 2006; **54**:628–634.
8. Bernatsky S, Hudson M, Suissa S. Anti-rheumatic drug use and risk of serious infections in rheumatoid arthritis. *Rheumatology* 2007; **46**:1157–1160.
9. Csajka C, Verotta D. Pharmacokinetic-pharmacodynamic modelling: history and perspectives. *Journal of Pharmacokinetics and Pharmacodynamics* 2006; **33**:227–279.
10. Herings RMC, Stricker BH, de Beor A, Bakker A, Sturmans F. Benzodiazepines and the risk of falling leading to femur fractures: dosage more important that elimination half-life. *Archives of Internal Medicine* 1995; **155**:1801–1807.
11. Petri M, Lakatta C, Magder L, Goldman D. Effect of prednisone and hydroxychloroquine on coronary artery disease risk factors in systemic lupus erythematosus: a longitudinal data analysis. *American Journal of Medicine* 1994; **96**:254–259.

12. Karp I, Abrahamowicz M, Fortin PR, Pilote L, Neville C, Pineau CA, Esdaile JM. Recent corticosteroid use and recent disease activity: independent determinants of coronary heart disease risk factors in systemic lupus erythematosus? *Arthritis & Rheumatism-Arthritis Care & Research* 2008; **59**:169–175.
13. Hauptmann M, Wellmann J, Lubin JH, Rosenberg PS, Kreienbrock L. Analysis of exposure-time-response relationships using a spline weight function. *Biometrics* 2000; **56**:1105–1108.
14. Vacek PA. Assessing the effect of intensity when exposure varies over time. *Statistics in Medicine* 1997; **16**:503–513.
15. Abrahamowicz M, Bartlett G, Tamblyn R, du Berger R. Modeling cumulative dose and exposure duration provided insights regarding the associations between benzodiazepines and injuries. *J Clin Epidemiol* 2006; **59**:393–403.
16. Sylvestre MP, Abrahamowicz M. Flexible modelling of the cumulative effects of time-dependent exposures on the hazard. *Statistics in Medicine* 2009; **28**:3437–3453.
17. Milliken C, Cooper SJ. Withdrawal symptoms from paroxetine. *Human Psychopharmacology-Clinical and Experimental* 1998; **13**:217–219.
18. Bespalov A, Jongen-Relo AL, Van Gaalen M, Harich S, Schoemaker H, Gross G. Habituation deficits induced by metabotropic glutamate receptors 2/3 receptor blockade in mice: reversal by antipsychotic drugs. *Journal of Pharmacology and Experimental Therapeutics* 2007; **320**:944–950.
19. MacKenzie T, Abrahamowicz M. Marginal and hazard ratio specific random data generation: applications to semi-parametric bootstrapping. *Statistics and Computing* 2002; **12**:245–252.
20. Abrahamowicz M, MacKenzie T, Esdaile JM. Time-dependent hazard ratio: modeling and hypothesis testing with application in lupus nephritis. *Journal of the American Statistical Association* 1996; **91**:1432–1439.
21. Sylvestre MP, Abrahamowicz M. Comparison of algorithms to generate event times conditional on time-dependent covariates. *Statistics in Medicine* 2008; **27**:2618–2634.
22. PermAlgo: Permutational algorith to generate event times conditional on a covariate matrix including time-dependent covariates. R package. Version 1.0; 2010.
23. Akaike H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 1974; **AC-19**: 716–723.
24. Volinsky CT, Raftery AE. Bayesian information criterion for censored survival models. *Biometrics* 2000; **56**:256–262.
25. Quantin C, Abrahamowicz M, Moreau T, Bartlett-Esquilant G, MacKenzie T, Tazi MA, Lalonde L, Faivre J. Variation over time of the effects of prognostic factors in a population based study of colon cancer: comparison of statistical models. *American Journal of Epidemiology* 1999; **150**:1188–1200.
26. Benedetti A, Abrahamowicz M, Goldberg MS. Accounting for data-dependent degrees of freedom selection when testing the effect of a continuous covariate in generalized additive models. *Communications in Statistics-Simulation and Computation* 2009; **38**:1115–1135.
27. R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria, 2010.
28. Altman DG, Andersen PK. Bootstrap investigation of the stability of a Cox Regression Model. *Statistics in Medicine* 1989; **8**:771–783.
29. Hastie TJ, Tibshirani RJ. *Generalized Additive Models*. Chapman & Hall/CRC: New York, 1990.
30. Neutel CI, Hirdes JP, Maxwell CJ, Patten SB. New evidence on benzodiazepine use and falls: the time factor. *Age & Ageing* 1996; **25**:273–278.
31. Canadian Pharmacists Association. Benzodiazepines. *Compendium of Pharmaceuticals and Specialties*. Ottawa, ON; 2005.
32. Yang YH. Can the strengths of AIC and BIC be shared? A conflict between model indentification and regression estimation. *Biometrika* 2005; **92**:937–950.
33. Ramsay JO. Monotone regression splines in action (with discussion). *Statistics in Science* 1988; **3**:425–461.
34. Abrahamowicz M, MacKenzie T. Joint estimation of time-dependent and non-linear effects of continuous covariates on survival. *Statistics in Medicine* 2007; **26**:392–408.
35. Hastie T, Tibshirani R. Exploring the nature of covariate effects in the proportional hazards model. *Biometrics* 1990; **46**:1005–1016.
36. Kooperberg C, Stone CJ, Truong YK. Hazard regression. *Journal of the American Statistical Association* 1995; **90**:78–94.
37. Automated data systems available for pharmacoepidemiology studies. In *Pharmacoepidemiology*, Strom BL (ed.). John Wiley & Sons Ltd: Chichester, 2010; 217–346.
38. Brookhart MA, Wang PS, Solomon DH, Schneeweiss S. Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable. *Epidemiology* 2006; **17**:268–275.
39. Abraham NS, Castillo DL, Hartman C. National mortality following upper gastrointestinal or cardiovascular events in older veterans with recent nonsteroidal anti-inflammatory drug use. *Alimentary Pharmacology & Therapeutics* 2008; **28**:97–106.
40. Walker AM. Confounding by indication. *Epidemiology* 1996; **7**:335–336.
41. McMahon AD. Approaches to combat with confounding by indication in observational studies of intended drug effects. *Pharmacoepidemiology and Drug Safety* 2003; **12**:551–558.
42. Brookhart MA, Rassen JA, Schneeweiss S. Instrumental variable methods in comparative safety and effectiveness research. *Pharmacoepidemiology and Drug Safety* 2010; **19**:537–554.
43. Abrahamowicz M, Tamblyn R. Drug utilization patterns. In *Encyclopedia of Biostatistics,* Armitage P, Colton T (eds), 2nd ed. John Wiley & Sons, Ltd: Chichester, 2005; 1533–1553.
44. Sun LQ, Zhou X. Inference in the additive risk model with time-varying covariates subject to measurement errors. *Statistics & Probability Letters* 2008; **78**:2559–2566.
45. Berhane K, Hauptmann M, Langholz B. Using tensor product splines in modeling exposure-time-response relationships: application to the Colorado Plateau Uranium Miners cohort. *Statistics in Medicine* 2008; **27**:5484–5496.
46. Richardson DB. Latency models for analyses of protracted exposures. *Epidemiology* 2009; **20**:395–399.