STATISTICS IN MEDICINE
Statist. Med. 201; **1**:1 *Prepared using* simauth.cls *[Version: 2002/09/18 v1.11]*

Supplementary material for Flexible modeling of the cumulative effects of time-dependent exposures on the hazard

Marie-Pierre Sylvestre[1] and Michal Abrahamowicz[1,*]

[1] *Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, 1020 Pine Avenue West, Montreal, Quebec, Canada, H3A 1A2*

The current document presents supplementary material that, for sake of brevity, could not be included in our paper *Flexible modeling of the cumulative effects of time-dependent exposures on the hazard*. Specifically, in this document, we describe the design of the simulation studies in enough details to permit replication of our results. In addition, we extend the analysis of the simulated results presented in the paper. Finally, this document offers a few insights on software specification for user interested in implementing the weighted cumulative exposure model presented in the paper. While some of the Equations and Figures referenced in this document refer to the paper, the articles cited here can be found at the end of the document.

1. Design of simulations and data generation

To validate our model and investigate its properties, we simulated a hypothetical prospective cohort study of the association between exposure to a single drug and time to an adverse event. The cohort consisted of 500 new users of the drug, and time zero was defined as the first day of the drug use [1]. Individuals could interrupt and resume their treatment repeatedly thereafter. The follow-up was limited to one year. In the next three sub-sections, we describe: (i) the generation of a matrix of individual drug use patterns, that was kept fixed across the simulations; (ii) the selection of alternative weight functions and parameters used in different data-generation scenarios; and (iii) the generation of event times conditional on the weighted cumulative doses.

*Correspondence to: McGill University Health Centre, 687 Pine avenue West, V building, Montreal, Quebec, Canada, H3A 1A1. Phone: (514) 934-1934 ext. 44712, Fax: (514) 934-8293, email: michal.abrahamowicz@mcgill.ca.

Contract/grant sponsor: M-P S. holds a Canadian Institutes for Health Research Doctoral Award. M.A. is a James McGill Professor at McGill University. This research was supported by grants from the National Sciences and Engineering Research Council of Canada (#228203) and the Canadian Institutes for Health Research (#MOP-8127).

1.1. Generation of time-dependent exposures

The drug treatment was assumed to vary in dose and duration both between individuals and over time within an individual. Since individuals could interrupt and then resume the treatment repeatedly, we generated consecutive periods of drug use, each followed by a period during which the treatment was interrupted. The duration, in weeks, of the initial treatment for individual i, was generated from a lognormal distribution with mean of $\log(0.5)$ and standard deviation of $\log(0.8)$, and rounded up to the next week. At the end of that week, the subject was assumed to start the first period of interruption, when the subject was assumed not to be exposed to the drug. The duration of the first interruption was generated from the same lognormal distribution. Then, the subsequent, alternating periods of use and interruptions were generated similarly, until the end of follow-up was reached. For each uninterrupted period of use, the standardized daily dose [1] was assumed to remain constant across the entire period. The daily dose $X(t)$ in equation (2), was randomly assigned the values of 0.5, 1, 1.5, 2, 2.5, or 3, with the recommended daily dose corresponding to 1.

1.2. Weight functions

We considered 6 different scenarios, each corresponding to a different true weight function, each defined over a $[0, 1]$ interval, corresponding to one year. As in (4), each weight function described how the relative weight of past exposure on the risk at current time u changed with increasing time $(u - t)$ elapsed since the exposure. Alternatively, each function can be interpreted as showing how the impact of current exposure changes with increasing time since exposure.

The six weight functions are shown as white curves in the six panels of Figure 1. The first two scenarios assumed that weights decreased monotonically as time since exposure increased. For scenario 1, we used an *exponential* decay function with $w(u-t) = 7e^{-7(u-t)}$ while for scenario 2, we used a *Bi-linear* function $w(u-t) = 1 - \frac{u-t}{50/365}$ for $u - t \leq \frac{50}{365}$ and 0 otherwise. Scenarios 3 and 4 corresponded to non-monotonic functions where the weights first increased and then decreased. The *Early peak* function of scenario 4 corresponded to the density of a $N[0.04; 0.05]$ distribution, while the *Inverted U* function for scenario 3, to the density function of a $N[0.2; 0.06]$ distribution, both left-truncated at $t = 0$. In comparison with the *Inverted U*, in the *Early peak* scenario, the maximum weight was assigned to more recent doses and the weights declined more sharply afterward. Scenario 5 corresponded to a *Constant* weight function with $w(u-t) = \frac{1}{180}$ for $0 \leq u - t \leq \frac{180}{365}$, so that the resulting WCE was in fact a standard un-weighted time-dependent cumulative dose variable $\sum X(t)$, calculated over the previous 6-month period. In the scenario 6, we considered a weight function that initially increased, reached a plateau at around $u - t = 180/365$ and then started to decrease to reach 0 at $u - t = \frac{240}{365}$. This function was labeled *Hat* and was specifically designed to investigate the impact of imposing an incorrect *a priori* constraint on the weight function. Indeed, in the analysis we constrained all weight functions to reach 0 at $u - t = \frac{180}{365}$ where the *Hat* function was still at its maximum value. To enhance comparability, all the six true weight functions were standardized so that the area under each function summed up to 1 over the interval $[u, u - 180/365]$.

1.3. Events generation

We generated event times conditional on the WCE, using the permutational algorithm, specifically developed and validated for simulating event times conditional on time-dependent covariates and/or effects [2, 3]. The permutational algorithm involves three major steps: (i) generating individual vectors of time-dependent covariate values; (ii) generating the event times, as well as random censoring times, from the pre-specified marginal distributions, independent of covariates; (iii) matching individual event times with individual covariate vectors based on pre-specified assumptions about the covariates impact on the hazard [4]. Matching at step (iii) is performed so that probability of a subject, who remained at risk until time t, with the time-dependent covariate vector $X_i(t)$, being matched with the event at time t is proportional to the subjects current hazard $h(t|X_i(t))$ [2, 3]. A detailed description of the algorithm and its validation can be found in [3].

Our implementation of the permutational algorithm to generate each simulated sample involved the following steps. First, for a given scenario with a pre-specified true weight function, we calculated the 500 individual vectors $WCE_i(u)$, of the true values of the time-dependent variable WCE in (2), for each day of potential follow-up $0 < u < 365$. To this end, for each subject $i = 1, .., 500$, we used the individual vector of the time-dependent daily drug exposure $X_i(t)$, $0 < t < 365$ days (generated in Section 1.1), and the corresponding weight function $w(ut)$ (as described in Section 1.2). We then generated $N = 500$ independent event times τ_i, independent of the exposure, assuming their marginal distribution is uniform $U[0, 365]$ days. We also generated $N = 500$ random right censoring times C_i, independent of exposure and event times, from uniform $U[0, 730]$ days distribution which was selected to obtain an approximately 50% rate of right random censoring. Next, we determined the individual observed follow-up time $T_i = \min(\tau_i, C_i)$ and censoring/event status at the end of his/her follow-up $\delta_i = I(T_i = \tau_i)$, which resulted in about 250 (un-censored) events per simulated sample.

The final step of the permutational algorithm involved matching each of the N individual event or censoring times T_i with one of the N time-dependent WCE vectors. To this end, we first ranked the observed times T_i , $i = 1, ..., 500$ in increasing order: $T_i - 1 < Ti < T_{i+1}$. We then proceeded iteratively from the earliest to the latest time, by matching each consecutive time with one of the available at risk WCE vectors, and then removing the matched WCE vector from all subsequent risk sets [4, 3]. If the observed time T_i corresponded to a censored observation ($\delta_i = 0$), we matched it with one among the WCE vectors in the corresponding risk set, using simple random sampling, with all at risk vectors assigned equal probabilities of being matched, independent of the individual exposure history. If, however, the observed time T_i corresponded to an un-censored event ($\delta_i = 1$), we employed weighted sampling, with probability of selecting a specific vector $WCE(u)$ vector proportional to the individual hazard, calculated at the event time Ti [4, 3]. Specifically, the probability of sampling a vector $WCE_s(T_i)$, from the corresponding risk set R_i , to be matched with the event at time T_i , was calculated as following:

$$P_s(T_i) = \frac{\exp(\ln(4))WCE_s(T_i)}{\sum_{p \in R_i} \exp(\ln(4))WCE_p(T_i)}. \tag{11}$$

Equation 11 assumes that that an unit increase in WCE corresponded to a 4-fold increase in the hazard (HR=4). Notice that because the weight function was standardized (see end

of Section 1.2), an unit increase in WCE corresponded to e.g. a difference between (a) an individual prescribed a standard dose of 1 for the entire half-year $[0 < t < \frac{180}{365}]$ window of etiologically relevant exposure, and (b) a subject not exposed at any time during this window.

1.4. Analysis of the simulated datasets

For each of the 1,000 samples simulated for a given scenario, we estimated unconstrained and constrained models with 1 to 3 interior knots, and used the BIC to select the best-fitting model. Regardless of the true shape of the weight function, we assumed that the user would select a six-month interval $[0, \frac{180}{365}]$ as the maximum window of potentially etiologically relevant exposures. Accordingly, we set $a = \frac{180}{365}$ for all spline models estimated in our simulations.

To assess the accuracy of our estimates of the exposure impact on the hazard, we simply reconstructed $\hat{\beta}$ as:

$$\hat{\beta} = \sum_{\tau=0}^{a} \sum_{j=1}^{m} \hat{\gamma}_j B_j(\tau). \tag{12}$$

Given (6) and (7), and the fact that each true weight function was defined so that

$$\int_0^a w(t)dt = 1, \tag{13}$$

the value of $\hat{\beta}$ should approximate the true β, i.e. the $\ln HR$ associated with a constant dose $X(t) = 1$ over the entire time window $[u - a, u]$. Based on the distribution of the estimated $\hat{\beta}$ across simulated samples, the relative bias $\frac{\bar{\hat{\beta}}-\beta}{\beta}$, and the standard deviation (SD) of the estimates were obtained from (12).

Next, we assessed the accuracy of the estimated weight functions. To this end, for each simulated sample, we first obtained the normalized version of the estimated weight function, which respected the unit integral constraint in (13):

$$\hat{w}^*(u - t) = \sum_{j=1}^{m} \frac{\hat{\gamma}_j}{\hat{\beta}} B_j(u - t) \tag{14}$$

where $\hat{\beta}$ was calculated from (12). We then plotted a random sample of 100 estimated normalized weight functions against the corresponding true weight functions to investigate the ability of the method to recover the shape of the true weight function used to generate the data.

We also estimated an alternative simpler model in which the exposure was expressed with a time-dependent cumulative (unweighted) dose $CUMDOSE = \sum_{t=u-a}^{u} X(t)$.

Finally, we used the simulation-based hypothesis testing procedure described in Section 2.5 to test the BIC-optimal WCE model in each of the simulated samples against the null hypothesis of no association. We used a LRT with a corrected cutoff α_0^* defined as the 5^{th} percentile of the empirical distribution of the conditional p-values obtained from performing a LRT on the BIC-optimal model in each of the 1,000 samples generated under the null hypothesis. We also assessed whether the unconstrained BIC-optimal WCE model provided a better fit to the data than the unweighted cumulative exposure model, by computing the proportion of simulated samples in which the difference in BIC between the two models was considered significant. To this end, we used a cutoff corresponding to the 5^th percentile of the

empirical distribution of the BIC differences between the optimal-BIC WCE model and the unweighted cumulative dose model, in the simulations in which the true weight function was the *Constant* function.

2. Simulation results

In this section, we report the results of the analysis of 1,000 simulated datasets for each of the six *true* weight functions described in Section 1.2.

2.1. Accuracy of the estimated weight functions

Figures 1 and 2 show a random sample of 100 estimated normalized weight functions obtained from, respectively, the unconstrained and constrained spline models. Each estimate corresponds to the BIC-optimal model for a given simulated sample. In each of the Figures' panels, the true weight function used to generate the data is plotted in white. In most scenarios, except for the *Hat* and the *Constant* functions (panels (e) and (f)), the majority of the estimated weight functions were able to capture the shape of the true weight functions, albeit with some variation in the amplitude of the curves. The *Hat* and the *Constant* scenarios are discussed in more details later on.

In the unconstrained models, the estimates show considerable over-fitting bias in the right tail of the plot, especially in scenarios when the exposures that occurred relatively long ago had little impact on the risk (Figure 1). Instability at the tails of the estimated function is a known feature of B-splines [5]. As Figure 2 shows, constraining the weight functions to smoothly go to zero at the end of the exposure time window considerably reduced the variation of the estimates, even if the window $[0, \frac{180}{365}]$ was selected *a priori* and was often much longer than the actual time window of relevant exposures.

2.2. Model selection

For each scenario, we used the BIC to select the overall best model among the unconstrained and constrained spline-based WCE models, and the un-weighted cumulative dose model, as described in Section 1.4. In addition, for each scenario, we assessed the proportion of simulated samples that yielded a statistically significant association between the exposure and the outcome using the likelihood ratio test with the corrected cutoff value $\alpha_0^* = 0.039$ (see Section 2.5). The 5^{th} percentile of the BIC difference between non-nested best fitting WCE model and unweighted cumulative dose model was 0.25. Based on this finding, we decided to simply set the corrected cutoff point for testing if the WCE model fits better than the unweighted model at 0, which resulted in a slightly conservative test. Accordingly, the differential weights in the BIC-optimal WCE model were considered to significantly improve the fit to data whenever this model had lower BIC than the unweighted cumulative model.

For scenarios 1-4, corresponding to the weight functions that decreased to zero near $u - t = 180/365$, the constrained WCE models had a better fit than unconstrained models in more than 90% of the simulated data sets. The number of knots selected for these models depended on the curvature of the true weight function. For monotone functions like the *Exponential* and the *Bi-linear* functions (Panels (a) and (b) in Figure 1), the BIC selected models with 1 interior knot in more than 90% of the samples. This reflected the fact that for

monotone functions, the additional flexibility provided by adding extra knots did not improve the fit enough to compensate for the penalty due to increasing the model dimension.

Models with 2 knots were selected more frequently for the two non-monotone weight functions that had a bump within the support interval $[u-180/365, u]$, namely the *Early peak*, and the *Inverted U*. This reflected the need for a greater flexibility in modeling functions that had a local extremum and one or two inflexion points within the support interval. Constrained models with two knots were selected as the best fit models in more than half of the simulated samples for the *Inverted U* scenario, while in the case of the *Early peak* weight function, more than two thirds of the models with the lowest BIC were the constrained models with 1 knot. The *Early peak* weight function had a narrow and sharp peak around $u - t = \frac{30}{365}$ which, in many samples, was not detected by the WCE method. Indeed, as Panels (c) of Figures 1 and 2 show, several estimated weight functions for the *Early peak* scenario were decreasing monotonically.

For Scenarios 1-4, the simpler non-spline based models were almost never selected as the best model, except for the unweighted cumulative dose model, which had the lowest BIC in 5.8% and 2.2% of the simulated samples out of 1,000 for the *Exponential* and *Inverted U* functions, respectively.

In scenarios 1-4, where the true weight functions decreased to zero within the $[u-180/365, u]$ interval, all of the LRT tests for both constrained and the unconstrained modeled correctly detected a statistically significant effect of the exposure. The unweighted cumulative dose model also yielded significant results in all samples for scenarios 1,2, and 4, and in 98% of samples for scenario 3.

In the scenario where the true model was the unweighted cumulative exposure (*Constant* weight), the correct model with the cumulative unweighted exposure variable (CUMDOSE) had the best fit in 913 out of 1,000 simulated samples. This indicated that the spline-based WCE models did not improve significantly the fit to data compared with the unweighted model $\sum X(t)$. This largely reduces any potential concerns about the overfit bias shown in the panel (e) of Figures 1 and 2. Based on the results of the BIC comparison, almost all the bumpy spline estimates would be rejected in favor of the time unweighted constant model.

Finally, in Scenario 6, the true weight function *Hat* increased until $u - t = 180/365$ and decreased thereafter. This example was specifically selected to assess the performance of the method in a difficult situation when the *a priori* selected support interval for the estimated weight function $[u - 180/365, u]$, was not wide enough to capture the entire time window of etiologically relevant exposure. As Panel (f) of Figure 2 shows, in this scenario, the constrained weight function estimates were forced to go to zero at $u-t = 180/365$, where the true function was in fact at its maximum. Accordingly, the constrained WCE models provided very biased estimates. On the other hand, in contrast to other scenarios, in Scenario 6, the unconstrained models, were performing better in approximating the true shape of the weight function, at least within the $[u-180/365, u]$ window (Figure (f)). Indeed, for scenario 6, the unconstrained WCE model was selected as the best-fitting spline model in 70% of the simulated samples.

The unconstrained WCE estimates shown in panel (f) of Figure 2 suggest that even those exposures near the right end of the support window have a marked impact on the current risk. Moreover, in this scenario, the shapes of almost all unconstrained weight functions suggest that exposures that occurred about half a year ago may be much more important than the more recent exposures. This should provide a strong suggestion that the support interval $[u - 180/365, u]$, *a priori* set by the hypothetical user, is actually much too short. In this

sense, these results suggest that if the true situation corresponded to our *Hat* scenario, then the use of alternative models would provide a strong evidence that the initially selected support interval is too short. We believe that, given such evidence, most users would re-analyse the data while using a considerably larger support interval. Therefore, we carried out additional sensitivity analyses, for scenario 6. Specifically, we modified the window of the support interval for the splines from the half year to the full year, and re-estimated all the models.

The resulting estimates were consistent with the *Hat* with the extended support interval $[u-1, u]$ function (figure not shown).In almost all ($\geq 99\%$) samples, the best-fitting model for the extended window fitted data much better than the best model for the initial window. This would provide the user with additional empirical evidence that the window of etiologically relevant exposure extends beyond half a year.

2.3. Estimated strength of the exposure impact

We also investigated whether the estimated the effect of the WCE on the risk of an event accurately reflected the true value of β in equation (3), i.e. the regression coefficient used to generate the data. To this effect, we compared the normalized estimates of β, obtained from (12), with the true value of β. The results are shown in the left-hand-side column of Table I, separately for the unconstrained and constrained models.

[Table 1 about here.]

As Table I shows, for Scenarios 1-4, the relative bias for $\hat{\beta}$ was relatively small, mostly below 5%, for both unconstrained and constrained models. For Scenarios 5 and 6, the bias for the unconstrained model was small (respectively -0.14% and 0.14%), while the bias for the constrained model was considerable (-18.18% and -43.36%). This was expected given the fact that these two scenarios corresponded to weight function that did not decrease to zero within the pre-selected support interval. Finally, in an additional scenario 7, where there was no association between exposure and the time to event, the bias was small for both the unconstrained (-2.4%) and the constrained model (-0.01%).

2.4. Limitation of the simulation study

In our simulations, the dose of each new prescription is generated independently of the previous dose, and exposure/dose history does not depend on covariates. Future simulations should assess our method under different, possibly more realistic, assumptions about the pattern and (lower) frequency of dose changes [6], e.g. by determining a new dose as (current dose + Δ), with a plausible distribution for Δ, and/or by assuming that both the initial dose and its consecutive changes may depend on some covariates.

2.5. Software specification

The algorithms for the data generation and for computing the time-dependent covariates in (6) were coded in R version 2.6.1. Then, model (7) was fit using the R procedure `coxph` with Efron method for tie handling [7].

Alternatively, users may want to implement our method using their favorite software. This would require initial decisions regarding choices of (a) the length a of the time window of etiologically relevant exposures (see Section 2.2), and (b) number of knots for the spline

model (4), which determines the number m of splines in the basis (Section 2.2). Then, the method can be implemented through three main steps. First, the cubic B-spline basis needs to be computed, using iterative formulas shown e.g. by de Boor [8]. Second, the artificial time-dependent cumulative variables $D_j(u)$ in (6) are calculated. Then, any software for Cox's model with time-dependent covariates can be employed to fit model (7). Finally, if required, equation (8) may be used to estimate hazard ratios corresponding to different exposure patterns.

It should be noted that computation of the artificial time-dependent cumulative variables $D_j(u)$ in (6) can be memory-intensive, especially for datasets with long follow-up and/or large number of subjects. Our current strategy for large data sets is to compute the $D_j(u)$ variables for every subject over the time window of interest, and to reduce the dimensionality of the data sets by removing rows that do not correspond to a unit of time during which any event occurred. In addition, for very large data sets, it might be advisable to split the original dataset into manageable subsets, compute the $D_j(u)$ variables over these subsets, and then collapse the subsets back into one dataset on which the models will be estimated.

REFERENCES

1. Tamblyn R, Abrahamowicz M, du Berger R, McLeod P, and 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**(2):233-41.
2. MacKenzie T, Abrahamowicz M. Marginal and hazard ratio specific random data generation: Applications to semi-parametric bootstrapping. *Statistics and Computing* 2002; **12**(3):245–252.
3. Sylvestre MP, Abrahamowicz M. Comparison of algorithms to generate event times conditional on time-dependent covariates. *Statistics in Medicine* 2008; **27**(14): 2618-2634.
4. Abrahamowicz M, MacKenzie T, Esdaile JM. Time-dependent hazard ratio: modeling and hypothesis testing with application in lupus nephritis. *JASA* 1996; **91**:1432-9.
5. Hastie TJ, Tibshirani RJ. *Generalized Additive Models* Chapman and Hall, 1990.
6. Bartlett G and Abrahamowicz M and Tamblyn R and Grad R and Capek R and du Berger R. Longitudinal patterns of new Benzodiazepine use in the elderly. *Pharmacoepidemiology and Drug Safety* 2004; **13**(10):669-682.
7. R Development Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing: Vienna, 2006.
8. de Boor C. *A practical guide to splines.* Berlin Heidelberg New York: Springer 1978.

Table I. Estimated β from 1,000 simulations for each of the weight function scenarios

		Unconstrained models			Constrained models		
Weight	True β	mean($\hat{\beta}$)	SD($\hat{\beta}$)	% Bias	mean($\hat{\beta}$)	SD($\hat{\beta}$)	% Bias
1 Exponential	1.386	1.346	0.328	-2.89%	1.317	0.269	-4.98%
2 Bi-linear	1.386	1.390	0.385	0.29%	1.372	0.299	-1.01%
3 Inverted U	1.386	1.425	0.352	2.81%	1.467	0.322	5.84%
4 Early peak	1.386	1.385	0.361	-0.07%	1.372	0.305	-0.01%
5 Constant	1.386	1.384	0.332	-0.14%	1.134	0.248	-18.18%
6 Hat	1.386	1.388	0.304	0.14%	0.785	0.287	-43.36%
7 Null	0	-0.024	0.322	-2.40%	-0.009	0.259	-0.099%