PHARMACOEPIDEMIOLOGY AND DRUG SAFETY 2015; **24**: 456–467
Published online 4 September 2014 in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/pds.3700

ORIGINAL REPORT

The impact of exposure model misspecification on signal detection in prospective pharmacovigilance†

Rolina D. van Gaalen[1], Michal Abrahamowicz[1,2]* and David L. Buckeridge[1]

[1]*Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montréal, Québec, Canada*
[2]*Division of Clinical Epidemiology, McGill University Health Centre, Montréal, Québec, Canada*

ABSTRACT

Purpose Pharmacovigilance monitors the safety of drugs after their approval and marketing. Timely detection of adverse effects is important. The true relationship between time-varying drug use and the adverse event risk is typically unknown. Yet, most current pharmacovigilance studies rely on arbitrarily chosen exposure metrics such as current exposure or use in the past 3 months. The authors used simulations to assess the impact of a misspecified exposure model on the timeliness of adverse effect detection.

Methods Prospective pharmacovigilance studies were simulated assuming different true relationships between time-varying drug use and the adverse event hazard. Simulated data were analyzed by fitting conventional parametric and more complex spline-based estimation models at multiple, pre-specified testing times. The 'signal' was generated on the basis of the corrected model-specific *p*-value selected to ensure a 5% probability of incorrectly rejecting the null hypothesis of no association.

Results Results indicated that use of an estimation model that diverged substantially from the true underlying association-reduced sensitivity and increased the time to detection of a clinically important association.

Conclusions Time to signal detection in pharmacovigilance may depend strongly on the method chosen to model the exposure. No single estimation model performed optimally across different simulated scenarios, suggesting the need for data-dependent criteria to select the model most appropriate for a given study. Copyright © 2014 John Wiley & Sons, Ltd.

KEY WORDS—pharmacovigilance; prospective surveillance; timeliness; model selection; signal detection; time-varying exposure; pharmacoepidemiology

Received 4 February 2014; Revised 23 June 2014; Accepted 23 July 2014

INTRODUCTION

Monitoring of drug safety must continue after marketing. Even large clinical trials lack the power and generalizability to capture the complete safety profile of a drug.[1,2] Pharmacovigilance aims to detect, assess, understand, and prevent adverse events (AE).[3] Signals of a drug/AE association can trigger further observational analyses and clinical trials, which can lead to changes in marketing and the possible withdrawal of a drug.[4,5] The timeliness with which such signals can be detected depends on the data available and the monitoring methods being used. Since the associations may be weak, large databases are necessary to ensure adequate statistical power,[6] and administrative population-based databases linked with electronic health records are a natural choice for post-marketing surveillance due to their size, rapid availability, and diverse population.[7,8] Traditional study designs and statistical methods used in pharmacoepidemiology have been adapted for prospective pharmacovigilance to monitor hypothesized associations between specific drug/AE pairs using observational databases.[9–12] Surveillance usually continues as data accumulate over time. At predetermined time intervals, a regression model is estimated using the available data to test the hypothesized association.[10,11] A signal is generated according to a pre-specified stopping rule, usually when there is sufficient evidence to reject the null hypothesis of no association, otherwise testing continues.[13]

*Correspondence to: M. Abrahamowicz, Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, 687 Pine Avenue West, V-Pavilion, Room V2.20A, Montreal, QC H3A 1A1, Canada. Email: michal.abrahamowicz@mcgill.ca

†**Prior presentations:** this manuscript has been presented in part at the International Conference on Pharmacoepidemiology and Therapeutic Risk Management in Montreal, QC, Canada (25–28 August 2013) and at the International Society for Disease Surveillance in New Orleans, LA, USA (12–13 December 2013).

One important and challenging methodological issue that has received relatively little attention in the pharmacovigilance literature is the representation of drug exposure in the statistical model used to test for an association. In pharmacovigilance, exposure is typically represented, with only a few exceptions,[14] by simple binary indicators of either current use or use within an arbitrarily defined time interval.[12,15,16] Simple exposure models ignore information on dose, duration, and timing of the exposure, all of which tend to vary considerably in post-marketing use[17,18] and affect the AE risk.[13,19–21] Moreover, the dose-response and the possible lagged or cumulative effects of exposure may differ substantially depending on the specific drug and AE.[20] For example, drug-induced anaphylaxis generally occurs immediately following the first exposure,[22] whereas osteoporosis may be a result of slower biological changes, possibly induced by long-term cumulative effects of past exposures.[23,24] Simple exposure measures are unable to capture these effects.[11]

From both a pharmacological perspective[20] and a statistical perspective,[25,26] accurate representation of the exposure history offers clear advantages. In retrospective pharmacoepidemiology studies, models that inaccurately represent the true effects of dose and treatment duration may substantially lower the power to detect a drug/outcome association compared with more appropriate models.[21,27,28] Therefore, whereas accurate modelling of time-varying drug exposure may also be expected to improve the accuracy of prospective pharmacovigilance, its effects on signal detection in different clinically relevant circumstances remain to be assessed.

In this study, we used simulations to investigate the impact of exposure model misspecification on the timeliness of signal detection in prospective pharmacovigilance under different assumptions regarding the association between time-varying drug use and the hazard of an AE.

METHODS

We used simulations to evaluate and compare the performance of alternative exposure models in pharmacovigilance under a range of plausible conditions. The simulations mimicked real-life prospective pharmacovigilance studies of a putative association between a specific drug and a particular AE. Design and implementation of simulations involved three broad components: (i) simulating individual time-varying drug exposure patterns; (ii) specifying the 'true' way in which the past and the current exposures were assumed to affect the AE hazard; and (iii) determining the surveillance methods, including models used to test the association and the stopping rule for signal detection.

Simulating drug exposure patterns

We generated hypothetical cohorts of new users of a drug with the unexposed time of the new users defining 'control periods'. To ensure realistic assumptions, exposure patterns and dosage changes were randomly sampled from population-based data on benzodiazepine use amongst older adult residents of Québec, Canada. Section A.1 of the Supplementary Materials describes the data in detail. Dosage was standardized using the defined daily dose for a given benzodiazepine.[29] Table 1 summarizes information on benzodiazepine exposure patterns in this cohort. To capture the dynamic and time-dependent nature of real-life exposure and pharmacovigilance, we simulated a prospective study with a maximum of five calendar years of follow-up by randomly assigning each individual a treatment start time within the first two years of follow-up.

Simulating adverse events

We relied on the aforementioned benzodiazepine users cohort to generate a realistic distribution of AE times.[30,31] We used a previously validated algorithm[32] to assign event times to individual subjects based on their simulated time-varying exposure patterns.

To assign event times to individuals with specific exposure histories, we used 12 alternative simulation models that assumed different associations between exposure and the time to AE. Detailed model descriptions are given in Table 2. Model (l) assumed no association and was used to assess the false positive rate and calibrate the stopping rule (explained in later text). Other simulation models included six 'conventional' simple parametric models, often used in pharmacoepidemiology, an additional parametric model that assumed a withdrawal effect, and four weighted cumulative exposure (WCE) models,[25,27] discussed in the next paragraph. Briefly, some models assumed that the current exposure had the strongest impact on the hazard, whereas others are allowed for longer duration of effect; some models assumed that the hazard depended on dosage, whereas doses were irrelevant in others.

To simulate complex associations, we used WCE models in which the risk depends on the cumulative effect of the past doses.[25,33,34] More precisely, the WCE(t) at time t is a time-dependent covariate,

DOI: 10.1002/pds

Table 1. Duration and periods of benzodiazepine exposure for 105 331 new users over 3 years of follow-up

Variable	Mean (SD)	Median (Q_1–Q_3)	Range
Observation time (days)	934.7 (312.6)	1095 (956–1095)	1–1095
Total duration of benzodiazepine use (days)	218.0 (287.7)	80 (30–280)	1–1095
Number of periods of uninterrupted benzodiazepine use	4.4 (4.6)	2 (1–6)	1–69
Duration of uninterrupted periods of benzodiazepine use (days)	50.0 (88.6)	30 (24–30)	1–1095
Duration of interruption between periods of benzodiazepine use (days)	76.8 (146.2)	22 (6–69)	1–1092
Defined daily dose during periods of benzodiazepine use (DDD)	0.46 (0.38)	0.4 (0.25–0.6)	0.0021–16.9

SD, standard deviation.

Table 2. Detailed descriptions of the alternative time-varying models used to simulate data linking drug exposure with adverse events

		Definition of the exposure model	
Exposure model		The hazard of an adverse event on day *t* depended on…	Example drug/adverse event pair*
Parametric models			
(a)	Current use	… whether a person was exposed on day *t*.	Penicillin and anaphylaxis[22]
(b)	Current dose	… the dose to which a person was exposed on day *t*.	Oral corticosteroids and fracture risk[55]
(c)	Use in the past *X* days	… whether a person was exposed within the *X* days prior to and including day *t*. In this paper, we set $X = 30$.	Idiosyncratic effects,[56] for example, antiepileptic drugs and delayed allergic hypersensitivity reactions[57,58]; however, the irrelevance of dose for idiosyncratic effects is debatable[59]
(d)	Cumulative dose in the past *X* days	… the cumulative dose to which a person was exposed within the *X* days prior to and including day *t*. In this paper, we set $X = 30$.	
(e)	Duration of use	… the number of days for which a person was exposed prior to and including day *t*.	Low dose (≤7.5 mg/day) prednisone and cataracts[60]
(f)	Total cumulative dose	… the total cumulative dose to which a person was exposed prior to and including day *t*.	Antiepileptic drugs and changes in bodyweight[57]
(g)	Withdrawal effect	… the total cumulative dose to which a person was exposed prior to and including day *t* and on whether a person discontinued treatment within the 7 days prior to and including day *t*. That is, the withdrawal effect was characterized by a sharp increase in risk as soon as a drug was discontinued.	Statins and subarachnoid hemorrhage[61]
WCE models			
(h)	Delayed effect	… historical exposures, with the effect of exposures increasing, reaching a peak and then decreasing over time. This weight function was modelled using a normal distribution that was centred on the number of days prior to day *t* that was expected to have the maximum impact on hazard on day *t*. In this paper, we centred the distribution on month 3. (Figure 1, panel (h).)	Prenatal exposure to antiepileptics and carcinogenic effects[57]
(i)	Decaying effect	… current exposures, with the effect of exposures decreasing over time. This weight function was modelled using a normal weight function that was truncated such that the maximum weight occurred at day *t*. (Figure 1, panel (i).)	Antiepileptic drugs and adverse psychiatric effects[57,62]; oral contraceptives and venous thrombosis[63]
(j)	Decaying and delayed effect	… more recent exposures and certain historical exposures. This weight function was modelled using a mixture of models (h) and (i). (Figure 1, panel (j).)	Glucocorticoids and serious infection[28]
(k)	Dual effect	The direction of the hazard of an adverse event on day *t* was dependent upon the timing of historical exposures: more recent exposures were assumed to be harmful, whereas more distant exposures were protective. (Figure 1, panel (k).)	Didanosine therapy and cardiovascular disease in HIV patients[35]
No association model			
(l)	No association	The hazard of an adverse event was not in any way associated with exposures.	

*Note: the relevance of binary models is that (i) many studies ignore dose; (ii) many databases have no information on dosage; and (iii) in some studies, there many be no variation in dose. The examples provided in this table reflect the associations established in the literature but not necessarily the *true* nature of the association, which may still be unknown.

DOI: 10.1002/pds

calculated as the weighted sum of the past doses up to time *t*, with weights that depend on the time elapsed since a given dose was taken.[21,25] We used four weight functions to represent alternative complex associations between drug exposure history and the hazard (Figure 1; Table 2). To consider both medium-term and long-term effects, we assumed that the period during which the exposure affects the future hazard was, respectively, 6 months for models (h)–(j) and up to 1.5 years for model (k). Previous analyses have revealed that cumulative effects of various drugs might extend to at least 1.5–2 years.[28,35]

We generated 300 cohorts of 3000 individuals for each of the 11 separate simulation models (a)–(k) (Table 2), and—to increase the precision in estimating the false positive rate—1000 cohorts from no-association model (l). Section A.2 of the Supplementary Materials provides details on the hazard ratio and cumulative incidence rate.

Methods for conducting prospective surveillance

Alternative estimation models. We assumed that the 'true' simulation model used for cohort generation was unknown. Therefore, for each simulated cohort, we evaluated prospective surveillance using several alternative exposure models, including six conventional parametric models (a)–(f) and two flexible WCE models, one that accounted for dosage (continuous WCE) and one that did not (binary WCE). For the WCE estimation models, the weight functions were estimated flexibly using cubic regression B-splines with two interior knots over the time interval consistent with the WCE simulation models (h)–(k), resulting in four degrees of freedom for the estimated exposure effect.[25]

Models (c)–(d) associated the current hazard with the past exposures during a specified time interval (Table 2). Our original estimation models (c)–(d) assumed an 'ideal case scenario', that is, they specified the same interval that was used in the corresponding simulation models (30 days). However, in practice, the exposure time window is often arbitrarily defined and unlikely to *exactly* match the true, unknown period. Therefore, we also estimated alternative versions of models (c)–(d) that misspecified the period of exposures associated with the current risk (details presented in Table 3).

In our main simulations, we assumed random censoring, no exposure misclassification, complete ascertainment of AEs, no unmeasured confounding, and knowledge of the exact event times. Sensitivity analyses, in which some of these assumptions were relaxed, are described at the end of the Methods section.

Stopping rule for signal detection. Prospective surveillance involved repeated analyses at up to 20 testing times, equally spaced at 3-month intervals. At each testing time, each alternative estimation model

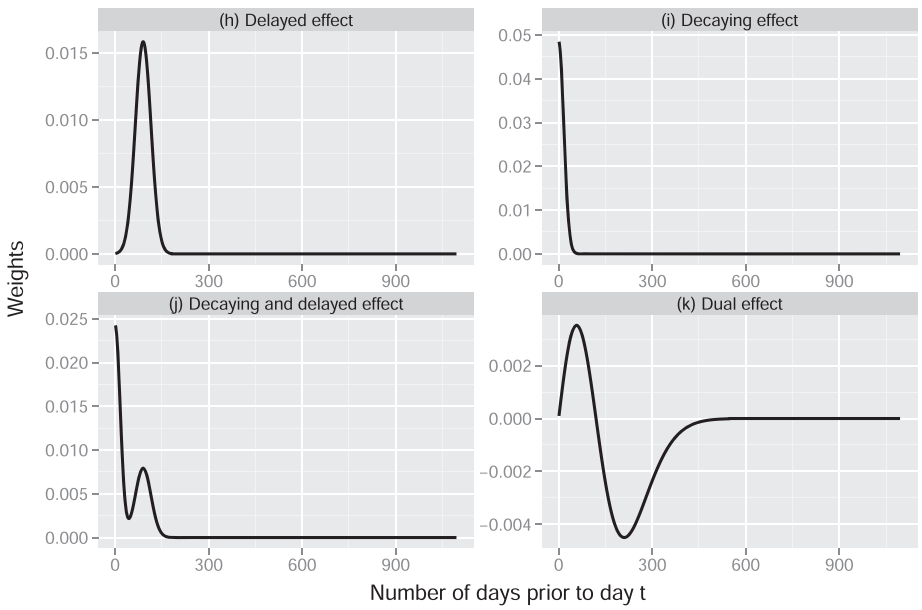


Figure 1. Graphical representation of the four weight functions used for the complex simulation models (h)–(k). The relative weight attributed to each past dose at day *t* (vertical axis) as a function of the number of days elapsed because the dose was taken

DOI: 10.1002/pds

Table 3. Detailed descriptions of the alternative time-varying estimation models that correctly specified and misspecified the true period during which past exposures affected the current risk

Exposure model		Definition of the exposure model
		The hazard of an adverse event on day *t* depended on…
Correctly specified parametric models for simulations models (c) and (d)		
(c)	Use in the past 30 days	… whether a person was exposed within the 30 days prior to and including day *t*.
(d)	Cumulative dose in the past 30 days	… the cumulative dose to which a person was exposed within the 30 days prior to and including day *t*.
Misspecified parametric models		
(c.1)	Use in the past 10 days	… whether a person was exposed within the 10 days prior to and including day *t*.
(c.2)	Use in the past 60 days	… whether a person was exposed within the 60 days prior to and including day *t*.
(d.1)	Cumulative dose in the past 10 days	… the cumulative dose to which a person was exposed within the 10 days prior to and including day *t*.
(d.2)	Cumulative dose in the past 60 days	… the cumulative dose to which a person was exposed within the 60 days prior to and including day *t*.
Correctly specified continuous and binary WCE models for models (h)–(k)		
		The weight functions were flexibly estimated over 6 months (180 days) for cohorts simulated using models (a)–(j) and 1.5 years (540 days) for cohorts simulated using model (k).
Misspecified continuous and binary WCE models		
		The weight functions were flexibly estimated over 1 year (365 days) for cohorts simulated using models (a)–(j) and 2 years (730 days) for cohorts simulated using model (k).
		The weight functions were flexibly estimated over 3 years (1095 days) for all cohorts.

WCE, weighted cumulative exposure.

used all data available until that time. Then, a model-specific likelihood ratio test statistic for the null hypothesis of no association was computed. For the spline-based WCE models, the number of degrees of freedom for the test was equal to four, corresponding to the number of spline coefficients.[25] The corresponding *p*-value for a given estimation model was used to 'decide', based on a predetermined threshold (discussed in later text), if sufficient evidence existed to detect a signal of a harmful association. For each simulated sample, we recorded the time to signal detection for each estimation model.

Obviously, basing signal detection on the uncorrected *p*-values would induce an inflated overall rate of false signals because of multiple testing.[36] In our context, the impact of multiple testing on false positive rate inflation is difficult to quantify analytically because it depends on the multivariate correlations between the values of the test statistic at consecutive testing intervals. Therefore, to control the expected overall false positive rate, we relied on simulations that assumed no association (Table 2, model (l)). For each of the 1000 cohorts simulated from model (l) and at each testing time, we retained the uncorrected *p*-value separately for each estimation model. Then, for different *p*-value cutoffs, we estimated the corresponding overall model-specific false positive rate as the proportion of simulated cohorts in which a given estimation model would generate a signal for at least one of the 20 testing times. For each estimation model, we defined the *p*-value that resulted in the observed 5% overall false positive rate as the corrected threshold for signal detection and used this cutoff across simulations with different 'true' data-generating models and at all testing intervals.

Methods and criteria to evaluate and compare model performance

We used Kaplan–Meier-like curves to compare the time to signal detection, across the eight estimation models, separately for each of the 11 different simulation models (a)–(k) (Table 2). For each estimation model, the Kaplan–Meier curve shows the proportion of cohorts that generated a signal up to, and including, a given testing interval. For each simulation and estimation model combination, the value at the 20th interval on the corresponding Kaplan–Meier curve estimates the final 'sensitivity' or the probability that a signal will be generated during the follow-up. We separately reported the median time to detection.[37,38]

To rank the estimation models with respect to the chronological order in which they detected the association, we used a rankogram.[39] The best-performing model has the highest probability at rank one, corresponding to the fastest signal detection, with probabilities monotonically decreasing for higher ranks. Since the performance of alternative estimation models depended strongly on whether the information on dose was relevant (in the simulation model) and taken into account by a given estimation model, we analyzed

DOI: 10.1002/pds

separately the results of the two types of simulations and ranked only those models that relied on the correct assumption regarding the role of doses.

Sensitivity analyses

We conducted additional simulations to assess the sensitivity of our results and conclusions to specific assumptions. First, we assessed the joint impact of (A) exposure measurement errors due to treatment non-adherence and (B) unmeasured confounding. Specifically, (A) we simulated different patterns of (i) occasional non-adherence resulting in the drug not being taken on some 'random' days or (ii) intentional permanent treatment discontinuation and allowed for variation of the intensity of non-adherence amongst subjects in both groups. In the same sensitivity analysis, (B) we simulated two time-varying unmeasured confounders, both associated with the current exposure and with lower AE risk.

The second sensitivity analyses assessed the generalizability of our results to more stable exposure patterns, with longer periods of use, and fewer and shorter treatment interruptions. Finally, we explored the impact of changing the chosen false positive rate and the strength of the underlying association. Section B of the Supplementary Materials provides further details of our simulation methods for all sensitivity analyses.

RESULTS

Different aspects of our results are presented in Figures 2 and 3 and Table 4. The six panels of Figure 2 use the Kaplan–Meier curves to compare the time to detection between alternative estimation models for the six selected simulated scenarios. Rankograms in Figure 3 present distributions of the ranks across relevant estimation models, in terms of timeliness of detection, for different simulated scenarios. Finally, Table 4 compares median times to detection across the estimation models.

To illustrate the impact of estimation model misspecification, we first interpret, as an example, the results for the data generated from the total cumulative dose model (Table 2, model (f)). As expected, the total cumulative dose model, consistent with the 'true' simulation model, offered the fastest signal detection (Figure 2f, dashed black curve) and was the only estimation model that detected a statistically significant association within the follow-up time in *all* simulated cohorts (Figure 2f). Therefore, the use of estimation models that misspecify the exposure effect decreases the sensitivity of signal detection, even by the 20th testing interval. Furthermore, most estimation models that ignored doses (Figure 2f, grey curves) had a substantially delayed detection of a harmful association compared with the models that accounted for dosage (black curves), and the median times to detection for the current or the past 30 days that use models (a) and (c) were double the corresponding median for the correct model (f) (Table 4, column (f)). Finally, among the four models that did account for dosages, the 'correct' model (f) was the first to generate a signal in 84% of the simulated cohorts (Figure 3f).

Results for the data generated from other simple parametric models (a)–(g) are generally consistent in

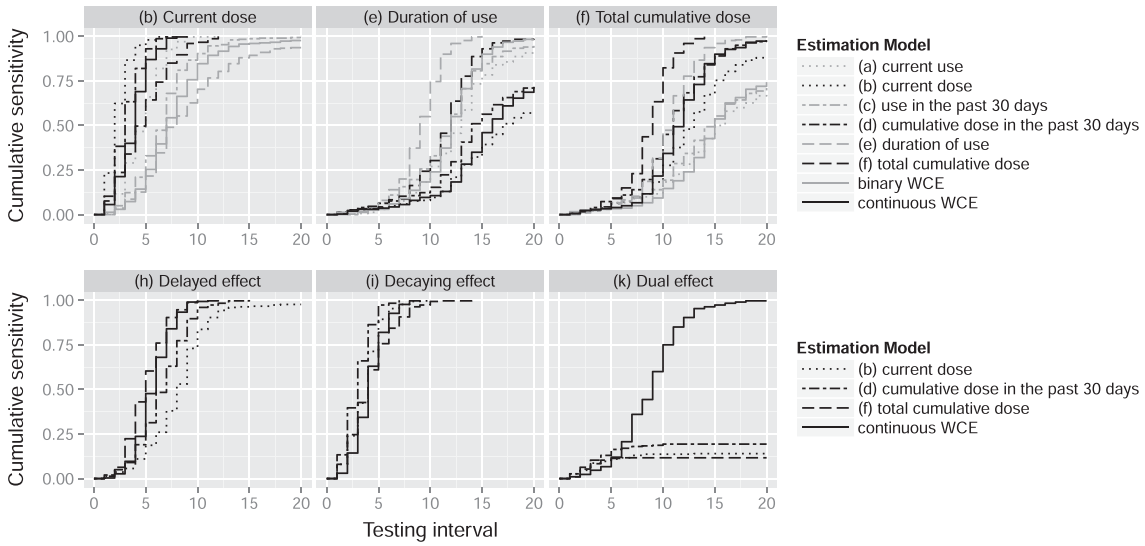
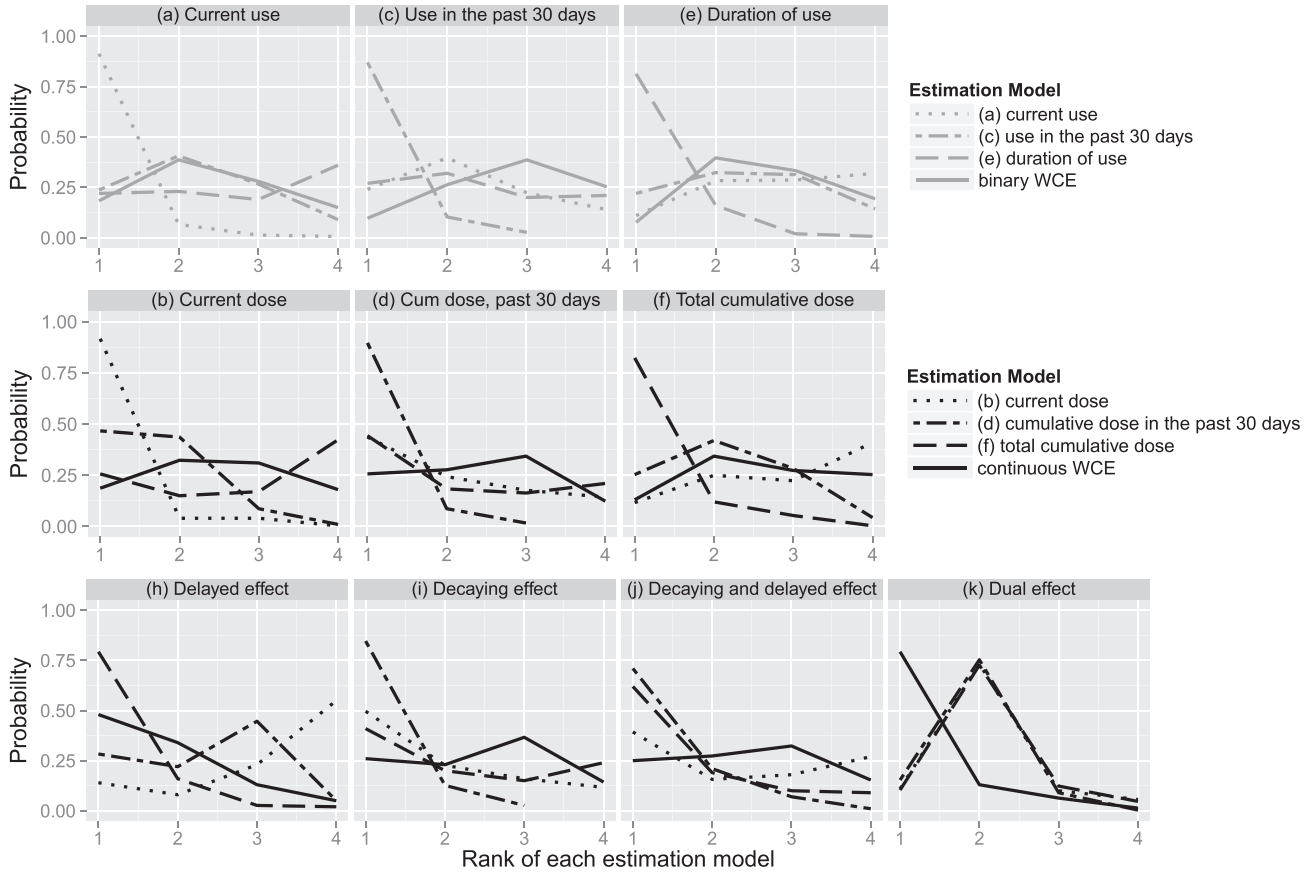
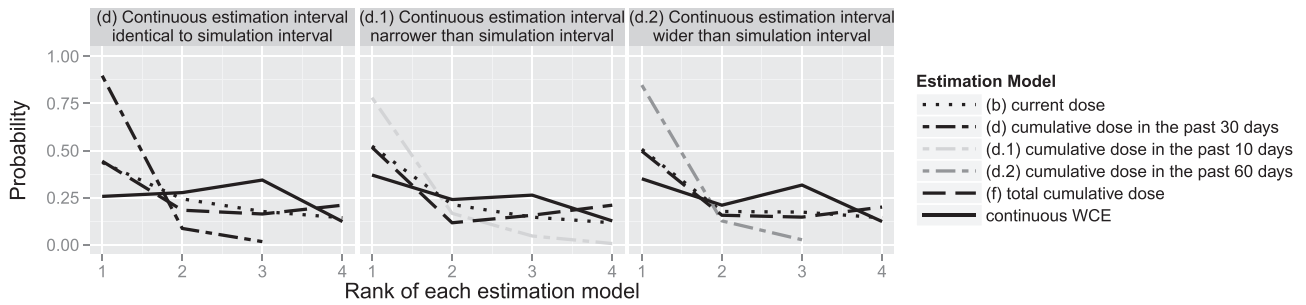


Figure 2. Comparison of times to signal detection generated by alternative estimation models for six selected simulation models. At each testing interval (horizontal axis), the curves indicate the proportion of the 300 cohorts generated for each simulation model that have already signalled an association (vertical axis)

DOI: 10.1002/pds



A) The estimation models shown in panels (a), (c), and (e) are only those for which risk did not depend on dosage, while those shown in panels (b), (d), (f), and (h) - (k) are only those for which risk did depend on dosage.



B) The estimation models shown in panels (d), (d.1), and (d.2) are only those for which risk depended on dosage. They compare the rank probabilities when the effect estimation interval of the cumulative dose estimation model is either identical to (panel (d)), narrower than (panel (d.1)), or wider than (panel (d.2)) that of the true cumulative dose simulation model (model (d)).

Figure 3. Comparison of estimation model rank probabilities in terms of timeliness of detection—relative to the other models included in each panel—(A) across 10 of the 11 simulation models and (B) between estimation intervals for the cumulative dose model (model (d)). Note that since multiple estimation models may have generated a signal at the same testing interval for a given simulated cohort (resulting in tied ranks), the probabilities at a given rank may not sum to 1

that misspecification of the estimation model may decrease the probability of detecting an adverse effect and considerably delay signal generation (Figures 2 and 3; Table 4). The impact of model misspecification increases if the estimation model assumes an exposure effect that diverges more substantially from the true association. For example, imposing an incorrect assumption regarding the presence (Figure 2, black

DOI: 10.1002/pds

Table 4. Median time to signal detection by simulation and estimation models

		Ever/never use simulation models		Duration of use simulation models		Cumulative dose simulation models			Weighted cumulative dose simulation models			
		(a) Current use	(c) Interval of use	(a) Current use	(e) Duration of use	(b) Current dose	(d) Interval dose	(f) Cumulative	(i) Decay	(h) Delay	(j) Decay and delay	(k) Dual effect
Estimation model	Period of increased risk (days)	1	30	1	1095	1	30	1095	180	180	180	550
Ever/never use												
(a) Current use	1	4	7	4	13.5	5	7	18	6	11	9	20
(c) Interval of use	10	5	6	5	13	6	7	17	6	11	8	20
	30	6	5	6	13	8	7	18	7	11	8	20
	60	7	6	7	12	8	8	16	8	9	8	20
Duration												
(a) Current use	1	4	7	4	13.5	5	7	18	6	11	9	20
(e) Duration of use	1095	7	7	7	10	9	8	11	8	8	7	20
Binary WCE												
	180 (i–j)/550 (k)	5	7	5	12	7	7	15	7	8	8	12
	365 (i–j)/730 (k)	6	7	6	12	8	8	14	7	9	8	13
	1095	6	7	6	11	9	8	13	8	9	8	13
Cumulative dose												
(b) Current dose	1	6	10	6	18	2	4	13	4	8	5	20
(d) Interval dose	10	6.5	9	6.5	17	3	4	12	3	8	5	20
	30	7	8	7	16	3	3	11	3	7	4	20
	60	7	8	7	14	4	3	11	3	6	4	20
(f) Cumulative	1095	10	10	10	12	5	4	9	4	5	4	20
Continuous WCE												
	180 (i–j)/550 (k)	9	11	9	16	4	4	12	4	6	5	9
	365 (i–j)/730 (k)	9	10	9	15	4	4	11	4	6	5	9
	1095	10	10.5	10	14	4	5	11	4	6	6	9

WCE, weighted cumulative exposure.

DOI: 10.1002/pds

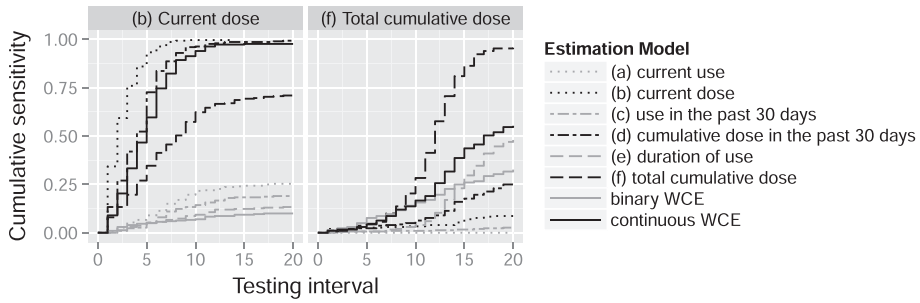


Figure 4. Comparison of times to signal detection generated by alternative estimation models for two selected simulation models in cohorts with simulated exposure measurement errors due to treatment non-adherence and unmeasured confounding. At each testing interval (horizontal axis), the curves indicate the proportion of the 300 cohorts generated for each simulation model that have already signalled an association (vertical axis)

curves) or the absence (Figure 2, grey curves) of a dose effect leads to systematically much longer detection times (Figure 2; Table 4).

Interestingly, when the data were generated from a WCE model (Table 2, models (h)–(j)), some simpler parametric estimation models still detected the harmful association faster than the correct flexible WCE model (Figures 3h–j). The power of WCE-based tests was reduced by the higher degrees of freedom required to model complex associations.[25] In contrast, parametric estimation models were often unable to detect the dual drug effect (Table 2, model (k)), even by the end of follow-up, whereas this effect was almost always detected by the flexible WCE model (Figure 2k).

Overall, although no single estimation model was uniformly 'optimal' across all simulated scenarios, neither the interval models (c)–(d) nor the WCE models were ranked as the worst models in any simulation, in terms of the longest median detection time (Figure 2) or the highest probability of (last) rank 4 (Figure 3). Indeed, in simulations that ignored dosage, the binary WCE model consistently ranked second best, being outperformed only by the 'correct' model (Figure 3a,c,e).

In general, regardless of the estimation model, in simulations where the true exposure effect depended on the total cumulative values of duration of the past use (Figure 2e) or the past doses (Figure 2f), signals were generated later than for shorter effects.

As expected, the performance of the estimation models also deteriorated if the time window over which the past exposure was assumed to affect current risk was misspecified (Table 4). Here, our results suggest that underestimating the length of the exposure time window may have a greater negative impact than overestimating it (e.g. Table 4, columns (d) versus (f)).

Figure 4 compares the times to signal detection for alternative estimation models in sensitivity analyses with both (i) exposure measurement errors due to non-adherence and (ii) unmeasured time-varying confounding. As expected, the joint impact of 'protective' unmeasured confounding and the exposure measurement errors generally delayed signal detection (Figure 4) compared with main simulations (Figure 2). However, these factors did not appear to affect the general trends noted earlier: estimation models consistent with the 'true' simulation models offered the fastest signal detection and the highest sensitivity of signal detection by the end of follow-up. Indeed, delays in detection were most marked for those estimation models that misspecified the underlying association (Figure 4). Additional sensitivity analyses further indicated that the general simulation results were robust with respect to variations in the stability of exposure patterns, the threshold for signal detection and the strength of the simulated association. Specifically, in all sensitivity analyses, the estimation model consistent with the true underlying association yielded the shortest time to signal detection and the highest sensitivity (Supplementary Materials, Section B).

DISCUSSION

Results of our simulations illustrate the impact of exposure model selection on the time to signal detection in pharmacovigilance. In most simulated scenarios, the adverse effects of a drug were detected most efficiently if the estimation model correctly specified the association used to generate the data. Use of estimation models that misspecified the true association could considerably reduce the probability of signal detection and induce important delays.

On the other hand, even when the true association involved complex cumulative but uniformly harmful effects, simple parametric models often generated signals more rapidly than the flexible WCE model. Although the latter would be essential to accurately

DOI: 10.1002/pds

describe the true cumulative effect,[21] the power of the corresponding test was reduced because of high degrees of freedom necessary to achieve modelling flexibility.[25] Therefore, as with other pharmacovigilance methods,[13] after initial signal detection, further analysis would be required to more accurately estimate and describe the nature and strength of the underlying association. However, the flexibility of the WCE model was instrumental in detecting complex 'dual' effects[28] of drugs or interventions associated, for example, with short-term risk increase, followed by a long-term protective effect (Figure 2k). An example of this type of association was reported between some antiretroviral drugs for HIV and cardiovascular risks.[35] Simple one-degree of freedom parametric models were often unable to detect signals in such complex situations (Figure 2k), because the average exposure effect over time would appear close to the null.

Our simulations suggest that without prior knowledge of the functional form of association, optimal pharmacovigilance is unlikely to be obtained by relying on an arbitrarily selected parametric model. Yet, methodology currently proposed for pharmacovigilance, such as those outlined by the Mini-Sentinel initiative, does not recommend which exposure models should be selected for analysis or on what criteria this selection should be based.[10] Applications of these methods,[11,12] and other sequential testing algorithms for pharmacovigilance,[40] involve simple exposure models such as a binary indicator of use/non-use or the total duration of the previous use that ignores the dosage and/or timing of exposure, without explaining the reason for the model choice. The current tendency to rely on arbitrarily chosen exposure models is reflected, for example, in the finding that several studies that investigated the same association, between use of glucocorticoids and the risk of infection in rheumatoid arthritis, employed a wide range of exposure models, including (i) the current use; (ii) the recent use, with different durations of the relevant time window; (iii) ever use; or (iv) the total cumulative (unweighted) dose.[41–46] Obviously, it is impossible that all these models correctly specify the true effect of glucocorticoids. Our finding that no single estimation model performed optimally across the range of simulated scenarios suggests the need to develop data-dependent criteria, which will help guide selection of the model(s) most appropriate for a given study.

Where information on dose is missing from the database, pharmacovigilance would need to rely on binary estimation models, which ignore the dosage. Where dose affects the time to an AE, using binary estimation models may delay signal detection and reduce 'sensitivity'. However, our results demonstrate that even in such analyses, accurately accounting for the duration and timing of the past drug use substantially improves the performance of binary estimation models (Figure 2b,f).

Our study has some limitations. First, our simulation models represent only a subset of the models that may represent the true relationships between different drug exposures and various AEs. However, most of the models that we considered represent associations that have been observed (Table 2), and that are conventionally tested for, in pharmacoepidemiology.

Second, similar to other studies in this area,[11,21,47] our main simulations assumed no unmeasured confounding and complete ascertainment of AEs. These simplifications helped us focus on the issues specific to exposure modelling. Clearly, different patterns of exposure measurement errors and/or unobserved confounding may be applied in different real-life pharmacovigilance studies. However, the fact that in our sensitivity analyses the general pattern of results was only slightly affected by a combination of frequent non-adherence and moderate unmeasured confounding suggests that our conclusions are robust to these factors. Yet, applications of our methods to real-life settings will require a careful consideration of these important issues.[7] Future research may also consider alternative designs of simulation studies.[48,49]

Third, we somewhat arbitrarily selected the false positive rate, chose our threshold for signal detection to be 'flat' (i.e. constant over time), and pre-set the number of testing intervals. In practice, where the number of testing points may be *a priori* unknown, it is more challenging to set pharmacovigilance thresholds.[50] The acceptable false discovery rate[51] may be selected to take into account both further investigations that will be required after signal detection and the AE severity.[11] Furthermore, the threshold may be either constant or decreasing over time[13,50] and may be determined on the basis of empirical studies with *a priori* known lack of an association.[52] In the current study, we implemented a 'generic' computationally expensive approach that relies on simulating the model-specific distribution of the test statistic based on multiple testing times that would be expected under the null hypothesis of no association. This allowed us to calibrate the p-value threshold to ensure the pre-specified overall false positive rate.[53,54] However, our approach requires specifying drug exposure patterns used in simulations. Future research should determine if 'standard' thresholds for use under a wide array of drug exposure patterns and distributions of the AEs under study can be identified.

DOI: 10.1002/pds

Prospective monitoring of large databases is necessary to ensure the timely detection of safety concerns related to pharmaceuticals. We have shown that time to detection may depend strongly on the method chosen to model the exposure. Since optimal pharmacovigilance is unlikely to be obtained by relying on only one arbitrarily selected parametric model, simultaneous testing of alternative models might prove to be effective. Future research should develop and evaluate analytical strategies that will combine efficient and timely signal detection with accurate control of the overall false discovery rate to account for both multiple testing and multiple estimation models.

CONFLICT OF INTEREST

Within the past 3 years, Rolina van Gaalen has worked as a part-time intern at Pfizer Canada. The work presented in this paper is neither related to her work at Pfizer nor funded by Pfizer.

KEY POINTS

- Time to signal detection in pharmacovigilance may depend strongly on the method chosen to model the exposure.
- The true form of the association between a time-varying drug exposure and the risk of an adverse effect is rarely known. Yet, using an estimation model that misspecifies the true underlying association may reduce sensitivity and increase the time to detection of a clinically important risk increase.
- No single estimation model performed optimally across different simulated scenarios.

ETHICS STATEMENT

The authors state that no ethical approval was needed.

ACKNOWLEDGEMENTS

This research was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) doctoral award and the Canadian Institutes for Health Research (CIHR) grants MOP 81275 and MOP 222481. Rolina van Gaalen is a doctoral candidate at McGill University. Michal Abrahamowicz is a James McGill Professor of Biostatistics at McGill University and David Buckeridge is the Canada Research Chair in Public Health Informatics. We thank the anonymous reviewers for their helpful suggestions.

REFERENCES

1. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". *Lancet* 2005; **365**: 82–93. doi:10.1016/S0140-6736(04)17670-8.
2. Black N. Why we need observational studies to evaluate the effectiveness of health care. *Br Med J* 1996; **312**: 1215–1218. doi:10.1136/bmj.312.7040.1215.
3. WHO. Medicines — Pharmacovigilance. WHO, Geneva, 2012. Available at: http://www.who.int/medicines/areas/qualitysafety/safetyefficacy/pharmvigi/en/ [12 September 2013].
4. FDA. Guidance drug safety information - FDA's communication to the public. U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research, 2007. Available at: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm [29 November 2013].
5. Briesacher BA, Soumerai SB, Zhang F, *et al.* A critical review of methods to evaluate the impact of FDA regulatory actions. *Pharmacoepidemiol Drug Saf* 2013; **22**: 986–994. doi:10.1002/pds.3480.
6. Kulldorff M, Davis RL, Kolczak M, *et al.* A maximized sequential probability ratio test for drug and vaccine safety surveillance. *Sequence Anal* 2011; **30**: 58–78. doi:10.1080/07474946.2011.539924.
7. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol* 2005; **58**: 323–337. doi:10.1016/j.jclinepi.2004.10.012.
8. Skegg DCG. Evaluating the safety of medicines, with particular reference to contraception. *Stat Med* 2001; **20**: 3557–3569. doi:10.1002/sim.1077.
9. Platt R, Carnahan RM, Brown JS, *et al.* The U.S. Food and Drug Administration's Mini-Sentinel program: status and direction. *Pharmacoepidemiol Drug Saf* 2012; **21**: 1–8. doi:10.1002/pds.2343.
10. Cook AJ, Tiwari RC, Wellman RD, *et al.* Statistical approaches to group sequential monitoring of postmarket safety surveillance data: current state of the art for use in the mini-sentinel pilot. *Pharmacoepidemiol Drug Saf* 2012; **21**: 72–81. doi:10.1002/pds.2320.
11. Gagne JJ, Rassen JA, Walker AM, *et al.* Active safety monitoring of new medical products using electronic healthcare data: selecting alerting rules. *Epidemiology* 2012; **23**: 238–246. doi:10.1097/EDE.0b013e3182459d7d.
12. Fireman B, Toh S, Butler MG, *et al.* A protocol for active surveillance of acute myocardial infarction in association with the use of a new antidiabetic pharmaceutical agent. *Pharmacoepidemiol Drug Saf* 2010; **21**: 282–290. doi:10.1002/pds.2337.
13. Nelson JC, Cook AJ, Onchee Y, *et al.* Methods for observational post-licensure medical product safety surveillance. *Stat Methods Med Res* 2011. doi:10.1177/0962280211413452.
14. Schuemie MJ. Methods for drug safety signal detection in longitudinal observational databases: LGPS and LEOPARD. *Pharmacoepidemiol Drug Saf* 2011; **20**: 292–299. doi:10.1002/pds.2051.
15. Kiri VA. A pathway to improved prospective observational post-authorization safety studies. *Drug Saf* 2012; **35**: 711–724. doi:10.2165/11630570-000000000-00000.
16. Harpaz R, DuMouchel W, Shah NH, *et al.* Novel data-mining methodologies for adverse drug event discovery and analysis. *Clin Pharmacol Ther* 2012; **91**: 1010–1021. doi:10.1038/clpt.2012.50.
17. Bartlett G, Abrahamowicz M, Grad R, *et al.* Association between risk factors for injurious falls and new benzodiazepine prescribing in elderly persons. *BMC Fam Pract* 2009; **10**: 1. doi:10.1186/1471-2296-10-1.
18. Christensen L, Sasané R, Hodgkins P, *et al.* Pharmacological treatment patterns among patients with attention-deficit/hyperactivity disorder: retrospective claims-based analysis of a managed care population. *Curr Med Res Opin* 2010; **26**: 977–989. doi:10.1185/03007991003673617.
19. McMahon AD, Evans JMM, McGilchrist MM, *et al.* Drug exposure risk windows and unexposed comparator groups for cohort studies in pharmacoepidemiology. *Pharmacoepidemiol Drug Saf* 1998; **7**: 275–280.
20. Csajka C, Verotta D. Pharmacokinetic-pharmacodynamic modelling: history and perspectives. *J Pharmacokinet Pharmacodyn* 2006; **33**: 227–279. doi:10.1007/s10928-005-9002-0.
21. Abrahamowicz M, Beauchamp ME, Sylvestre MP. Comparison of alternative models for linking drug exposure with adverse effects. *Stat Med* 2012; **31**: 1014–1030. doi:10.1002/sim.4343.
22. Neugut AI, Ghatak AT, Miller RL. Anaphylaxis in the United States: an investigation into its epidemiology. *Arch Intern Med* 2001; **161**: 15–21. doi:10.1001/archinte.161.1.15.

23. Mazziotti G, Canalis E, Giustina A. Drug-induced osteoporosis: mechanisms and clinical implications. *Am J Med* 2010; **123**: 877–884. doi:10.1016/j.amjmed.2010.02.028.
24. Grossman JM, Gordon R, Ranganath VK, *et al.* American college of rheumatology 2010 recommendations for the prevention and treatment of glucocorticoid-induced osteoporosis. *Arthritis Care Res* 2010; **62**: 1515–1526. doi:10.1002/acr.20295.
25. Sylvestre MP, Abrahamowicz M. Flexible modeling of the cumulative effects of time-dependent exposures on the hazard. *Stat Med* 2009; **28**: 3437–3453. doi:10.1002/sim.3701.
26. Abrahamowicz M, Tamblyn R. Drug Utilization Patterns. In Encyclopedia of Biostatistics, Armitage P, Colton T (eds.). John Wiley & Sons, 2005. doi:10.1002/0470011815.b2a04017.
27. Abrahamowicz M, Bartlett G, Tamblyn R, *et al.* Modeling cumulative dose and exposure duration provided insights regarding the associations between benzodiazepines and injuries. *J Clin Epidemiol* 2006; **59**: 393–403. doi:10.1016/j.jclinepi.2005.01.021.
28. Dixon WG, Abrahamowicz M, Beauchamp ME, *et al.* Immediate and delayed impact of oral glucocorticoid therapy on risk of serious infection in older patients with rheumatoid arthritis: a nested case-control analysis. *Ann Rheum Dis* 2012; **71**: 1128–1133. doi:10.1136/annrheumdis-2011-200702.
29. WHO. International language for drug utilization research. WHO, Geneva, 2013. Available at: http://www.whocc.no [10 September 2013].
30. Tamblyn R, Abrahamowicz M, du Berger R, *et al.* A 5-year prospective assessment of the risk associated with individual benzodiazepines and doses in new elderly users. *J Am Geriatr Soc* 2005; **53**: 233–241. doi:10.1111/j.1532-5415.2005.53108.x.
31. Cumming RG, Le Conteur DG. Benzodiazepines and risk of hip fracture in older people: a review of the evidence. *CNS Drugs* 2003; **17**: 825–837. doi:10.2165/00023210-200317110-00004.
32. Sylvestre MP, Abrahamowicz M. Comparison of algorithms to generate event times conditional on time-dependent covariates. *Stat Med* 2008; **27**: 2618–2634. doi:10.1002/sim.3092.
33. Breslow NE, Lubin JH, Marek P, *et al.* Multiplicative models and cohort analysis. *J Am Stat Assoc* 1983; **78**: 1–12.
34. Thomas D. Models for exposure-time-response relationships with applications to cancer epidemiology. *Annu Rev Public Health* 1988; **9**: 451–482. doi:10.1146/annurev.pu.09.050188.002315.
35. Xiao Y, Abrahamowicz M, Moodie E, *et al.* Flexible marginal structural models for estimating the cumulative effect of a time-dependent treatment on the hazard: Reassessing the cardiovascular risks of didanosine treatment in the Swiss HIV cohort study. *J Am Stat Assoc* in press. doi:10.1080/01621459.2013.872650.
36. Kulldorff M. Prospective time periodic geographical disease surveillance using a scan statistic. *J R Stat Soc Ser A Stat Soc* 2001; **164**: 61–72. doi:10.1111/1467-985X.00186.
37. Siegrist D, Pavlin J. Bio-ALIRT biosurveillance detection algorithm evaluation. *MMWR Morb Mortal Wkly Rep* 2004; **53**: 152–158.
38. Buckeridge DL, Burkom H, Campbell M, *et al.* Algorithms for rapid outbreak detection: a research synthesis. *J Biomed Inform* 2005; **38**: 99–113. doi:10.1016/j.jbi.2004.11.007.
39. Comparing multiple interventions methods group. The Cochrane Collaboration, 2013. Available at: http://cmimg.cochrane.org/node/109 [30 November 2013].
40. Brown JS, Kulldorff M, Chan KA, *et al.* Early detection of adverse drug events within population-based health networks: application of sequential testing methods. *Pharmacoepidemiol Drug Saf* 2007; **16**: 1275–1284. doi:10.1002/pds.1509.
41. Franklin J, Lunt M, Symmons D, *et al.* Risk and predictors of infection leading to hospitalisation in a large primary-care-derived cohort of patients with inflammatory polyarthritis. *Ann Rheum Dis* 2007; **66**: 308–312. doi:10.1136/ard.2006.057265.
42. Lacaille D, Guh DP, Abrahamowicz M, *et al.* Use of nonbiologic disease-modifying antirheumatic drugs and risk of infection in patients with rheumatoid arthritis. *Arthritis Rheum* 2008; **59**: 1074–1081. doi:10.1002/art.2391.
43. Smitten AL, Choi HK, Hochberg MC, *et al.*. The risk of hospitalized infection in patients with rheumatoid arthritis. *J Rheumatol* 2008; **35**: 387–393.
44. Schneeweiss S, Setoguchi S, Weinblatt ME, *et al.* Anti-tumor necrosis factor alpha therapy and the risk of serious bacterial infections in elderly patients with rheumatoid arthritis. *Arthritis Rheum* 2007; **56**: 1754–1764. doi:10.1002/art.22600.
45. Bernatsky S, Hudson M, Suissa S. Anti-rheumatic drug use and risk of serious infections in rheumatoid arthritis. *Rheumatology* 2007; **46**: 1157–1160. doi:10.1093/rheumatology/kem076.
46. Saag KG, Koehnke R, Caldwell JR, *et al.* Low dose long-term corticosteroid therapy in rheumatoid arthritis: an analysis of serious adverse events. *Am J Med* 1994; **96**: 115–123. doi:10.1016/0002-9343(94)90131-7.
47. Setoguchi S, Schneeweiss S, Brookhart MA, *et al.* Evaluating uses of data mining techniques in propensity score estimation: a simulation study. *Pharmacoepidemiol Drug Saf* 2008; **17**: 546–555. doi:10.1002/pds.1555.
48. Murray RE, Ryan PB, Reisinger SJ. Design and validation of a data simulation model for longitudinal healthcare data. *AMIA Annu Symp Proc* 2011; **2011**: 1176–1185.
49. Franklin JM, Schneeweiss S, Polinski JM, *et al.* Plasmode simulation for the evaluation of pharmacoepidemiologic methods in complex healthcare databases. *Comput Stat Data Anal* 2014; **72**: 219–226. doi:10.1016/j.csda.2013.10.018.
50. Nelson JC, Cook AJ, Yu O, *et al.* Challenges in the design and analysis of sequentially monitored postmarked safety surveillance evaluations using electronic observational health care data. *Pharmacoepidemiol Drug Saf* 2012; **21**: 62–71. doi:10.1002/pds.2324.
51. Storey JD, Tibshirani R. Statistical significance for genome wide studies. *Proc Natl Acad Sci U S A* 2003; **100**: 9440–9445. doi:10.1073/pnas.1530509100.
52. Ryan PB, Schuemie MJ, Welebob E, *et al.* Defining a reference set to support methodological research in drug safety. *Drug Saf* 2013; **36**: S33–S47. doi:10.1007/s40264-013-0097-8.
53. Mahmud M, Abrahamowicz M, Leffondré K, *et al.* Selecting the optimal transformation of a continuous covariate in Cox's regression: implications for hypothesis testing. *Commun Stat* 2006; **35**: 27–45. doi:10.1080/03610910500415928.
54. 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. *Commun Stat* 2009; **38**: 1115–1135. doi:10.1080/03610910902796032.
55. Van Staa TP, Leufkens HGM, Abenhaim L, *et al.* Use of oral corticosteroids and risk of fractures. *J Bone Miner Res* 2000; **15**: 993–1000. doi:10.1359/jbmr.2000.15.6.993.
56. Zaccara G, Franciotta D, Perucca E. Idiosyncratic adverse reactions to antiepileptic drugs. *Epilepsia* 2007; **48**: 1223–1244. doi:10.1111/j.1528-1167.2007.01041.x.
57. Perucca P, Gilliam FG. Adverse effects of antiepileptic drugs. *Lancet Neurol* 2012; **11**: 792–802. doi:10.1016/S1474-4422(12)70153-9.
58. Basedow S, Eigelshoven S, Homey B. Immediate and delayed hypersensitivity to corticosteroids. *JDDG* 2011; **9**: 885–888. doi:10.1111/j.1610-0387.2011.07718.x.
59. Chalasani N, Björnsson E. Risk factors for idiosyncratic drug-induced liver injury. *J Gastroenterol* 2010; **138**: 2246–2259. doi:10.1053/j.gastro.2010.04.001.
60. Curtis JR, Westfall AO, Allison J, *et al.* Population-based assessment of adverse events associated with long-term glucocorticoid use. *Arthritis Rheum* 2006; **55**: 420–426. doi:10.1002/art.21984.
61. Risselada R, Straatman H, van Kooten F, *et al.* Withdrawal of statins and risk of subarachnoid hemorrhage. *Stroke* 2009; **40**: 2887–2892. doi:10.1161/?STROKEAHA.109.552760.
62. Weintraub D, Buchsbaum R, Resor SR, *et al.*. Psychiatric and behavioural side effects of the newer antiepileptic drugs in adults with epilepsy. *Epilepsy Behav* 2007; **10**: 105–110. doi:10.1016/j.yebeh.2006.08.008.
63. Van Hylckama Vlieg A, Helmerhorst FM, Vandenbroucke JP, *et al.* The venous thrombotic risk of oral contraceptives, effects of oestrogen dose and progestogen type: results of the MEGA case-control study. *BMJ* 2009; **339**: b2921. doi:10.1136/bmj.b2921.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web site.