Drug Saf (2017) 40:1119–1129
DOI 10.1007/s40264-017-0555-9

ORIGINAL RESEARCH ARTICLE

Using Multiple Pharmacovigilance Models Improves the Timeliness of Signal Detection in Simulated Prospective Surveillance

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

Published online: 29 June 2017

Abstract

Introduction Prospective pharmacovigilance aims to rapidly detect safety concerns related to medical products. The exposure model selected for pharmacovigilance impacts the timeliness of signal detection. However, in most real-life pharmacovigilance studies, little is known about which model correctly represents the association and there is no evidence to guide the selection of an exposure model. Different exposure models reflect different aspects of exposure history, and their relevance varies across studies. Therefore, one potential solution is to apply several alternative exposure models simultaneously, with each model assuming a different exposure–risk association, and then combine the model results.

Methods We simulated alternative clinically plausible associations between time-varying drug exposure and the hazard of an adverse event. Prospective surveillance was conducted on the simulated data by estimating parametric and semi-parametric exposure–risk models at multiple times during follow-up. For each model separately, and using combined evidence from different subsets of models, we compared the time to signal detection.

Results Timely detection across the simulated associations was obtained by fitting a set of pharmacovigilance models. This set included alternative parametric models that assumed different exposure–risk associations and flexible models that made no assumptions regarding the form/shape of the association. Times to detection generated using a simple combination of evidence from multiple models were comparable to those observed under the ideal, but unrealistic, scenario where pharmacovigilance relied on the single 'true' model used for data generation.

Conclusions Simulation results indicate that, if the true model is *not* known, an association can be detected in a more timely manner by first fitting a carefully selected set of exposure–risk models and then generating a signal as soon as any of the models considered yields a test statistic value below a predetermined testing threshold.

Key Points

Combining the results from alternative exposure–risk models applied simultaneously can detect unknown associations between drugs and adverse events in a timely manner.

To detect unknown associations in pharmacovigilance, we recommend combining evidence from a carefully selected set of exposure–risk models and generating a signal as soon as any of the models considered yields a test statistic value below a predetermined testing threshold.

Electronic supplementary material The online version of this article (doi:10.1007/s40264-017-0555-9) contains supplementary material, which is available to authorized users.

✉ Rolina D. van Gaalen
rolina.vangaalen@mail.mcgill.ca

[1] Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, QC, Canada

[2] Division of Clinical Epidemiology, McGill University Health Centre, Montreal, QC, Canada

1 Introduction

Pharmacovigilance monitors the safety of a drug. The importance of efficient and accurate adverse event (AE) detection as a means for reducing adverse health outcomes has stimulated research to improve the methods used for pharmacovigilance [1–4].

One important challenge in real-life studies of drug effects is the need to correctly specify the exposure–risk model, which should account for variability in the dose, duration, and timing of drug exposures [5, 6]. Each of these components of time-varying exposure may affect AE risks [7–11], but they are often ignored in pharmacovigilance studies, which typically rely on very simple, arbitrarily selected exposure models [12] such as ever/never use or current use [13, 14]. For most pharmacovigilance studies there is insufficient prior knowledge about the way that drug exposure may affect the risks of the AE of interest, and this uncertainty hinders the ability of a researcher to identify a priori the single 'etiologically correct' exposure model.

A consequence of this methodological challenge is the wide array of exposure models used in different studies of the same association. For example, several pharmacoepidemiological studies that investigated the association between the use of non-steroidal anti-inflammatory drugs (in prevalent and/or new users) and the risk of myocardial infarction relied on a wide range of exposure models, including (1) current use or dose (low/moderate or high), with definitions of 'current' exposure ranging from exposure on the same day as the index date to exposure 30 days prior to the index date; (2) recent use, with different durations of the relevant time window; (3) ever use; and (4) duration of use, provided the study participant was a current user [15–20]. Another example, reported in Tournier et al. [21], concerned comparison of published studies that assessed the association between exposure to different mood stabilizers and the risk of metabolic events. Time-varying exposure was defined differently across studies, which could partly explain some conflicting results. For example, when exposure was defined as total duration of past use, the risk of diabetes mellitus did not vary across users of different mood-stabilizers [22]. In contrast, when binary indicators of exposure to a specific drug were used, certain mood stabilizers were found to be associated with a significantly higher risk of diabetes than others [23].

Current pharmacovigilance guidelines offer no advice on how to select an exposure model [24] and many applications have relied on simple exposure models, without explaining the reasons for having chosen a specific model (e.g., [13, 25, 26]). Yet, very different exposure models may be necessary depending on the type of AE (e.g., [27]) and pharmacokinetics/pharmacodynamics of the drug of interest (e.g., [28, 29]).

While the first aim of pharmacovigilance is signal detection, and a complete understanding of the underlying association is not expected before a more detailed investigation has been conducted, misspecification of the true exposure–risk association in a model used for pharmacovigilance can increase the time to signal detection [30] and substantially reduce the statistical power [7], and, therefore, can decrease the probability of detection [7, 30]. Since different exposure models are sensitive to different aspects of exposure history [31–33], one potential solution is to apply several alternative exposure models simultaneously, each assuming a different exposure–risk association, and then use appropriate statistical methods to combine their results, as done in other fields [34–37]. However, to the best of our knowledge, neither the analytical challenges of combining results from alternative exposure models nor the potential gains from this approach have been investigated systematically in the context of pharmacovigilance. In this article, we use simulation to investigate the potential benefits of combining results of multiple exposure–risk models. We compare the resulting times to signal detection, under a range of simulated, clinically plausible associations between a time-varying drug exposure and the hazard of an AE.

2 Methods

2.1 Simulations: Design, Assumptions, and Data Generation

We simulated cohorts of 3000 new users of a hypothetical drug, with individual exposures beginning at different times and 5 years of follow-up for each person. Data generation involved three consecutive steps. We first randomly sampled 3000 time-varying dosage patterns from actual data on the dispensing of benzodiazepines to adults over the age of 65 years living in the Canadian province of Québec. Second, we sampled 3000 AE times, independent of the previous exposure pattern sampling, using times to hospitalization for a fall-related injury observed in the same cohort (details in Electronic Supplementary Material 1). In the third step, each time-varying exposure pattern generated in step 1 was matched with one of the event/censoring times generated in step 2. Matching was performed in a way that ensured that simulated data were consistent with a prespecified 'true' exposure–risk model [38, 39]. We relied on a previously validated 'permutational algorithm' that matches an individual time-varying exposure pattern with an AE at time t, using weighted

sampling from the corresponding risk set $R(t)$, with weights proportional to the individual hazard at time t [38, 39]. Figure 2 in Electronic Supplementary Material 1 illustrates this simulation technique graphically.

These three steps were used to simulate, in separate 'simulation scenarios', data from 11 alternative 'true' exposure models, each representing a different plausible association between time-varying exposure to a drug and the time to an AE. Simulation scenarios included six simple models, (a)–(f) in Table 1, used routinely in pharmacoepidemiology and one model, (g), that assumed a withdrawal effect. Four additional simulation scenarios accounted for more complex effects of past exposures, including delayed or slowly decaying effects. Data for these models, (h)–(k) in Table 2, were generated using weighted cumulative effect (WCE) models [40, 41], where the AE hazard on day t depended on the weighted sum of the past doses up to day t. Weights reflected the relative importance of doses taken at different times in the past [40, 41]. Each of the four WCE models assumed a different weight function (see Figure 3 in Electronic Supplementary Material 1).

We generated 1000 cohorts of 3000 new users for each of scenarios (a)–(k), corresponding to the

Table 1 Detailed descriptions of the alternative time-varying parametric models used to simulate, and then (re-)analyze, data linking drug exposure with adverse events

Exposure model		Exposures associated with hazard of an adverse event on day t	Simulation scenario (S) and/or pharmacovigilance model (P)[a]		Subsets containing each pharmacovigilance model[b]					
(a)	Current use	Use on day t	S	P	(i)	(iii)	(iv)	(vi)	(viii)	
(b)	Current dose	Dosage on day t	S	P	(i)	(iii)	(v)	(vi)		(ix)
(c)	Use in the past 30 days	Use within the 30 days prior to and including day t	S	P	(i)	(iii)	(iv)	(vi)	(viii)	
(c.1)	Use in the past 10 days	Use within the 10 days prior to and including day t		P				(vi)	(viii)	
(c.2)	Use in the past 60 days	Use within the 60 days prior to and including day t		P				(vi)	(viii)	
(d)	Cumulative dose in the past 30 days	Cumulative dose within the 30 days prior to and including day t	S	P	(i)	(iii)	(v)	(vi)		(ix)
(d.1)	Cumulative dose in the past 10 days	Cumulative dose within the 10 days prior to and including day t		P				(vi)		(ix)
(d.2)	Cumulative dose in the past 60 days	Cumulative dose within the 60 days prior to and including day t		P				(vi)		(ix)
(e)	Duration of use	Number of days of use prior to and including day t	S	P	(i)	(iii)	(iv)	(vi)	(viii)	
(f)	Total cumulative dose	Total cumulative dose prior to and including day t	S	P	(i)	(iii)	(v)	(vi)		(ix)
(g)	Withdrawal effect	Total cumulative dose prior to and including day t, and discontinuation of treatment within the 7 days prior to and including day t	S		–					
(l)	No association	The hazard of an adverse event was not in any way associated with exposures	S		–					

[a] Exposure models used to simulate data are indicated with an 'S' (Sect. 2.1). Exposure models used to analyze the data either as single models or as part of a subset of models are indicated with a 'P' (Sect. 2.2)

[b] The pharmacovigilance models included in each pharmacovigilance subset, (i)–(ix), are indicated in this column (Sect. 2.2)

Table 2 Detailed descriptions of the alternative time-varying flexible weighted cumulative effect models and model assuming no association used to simulate or (re-)estimate data linking drug exposure with adverse events

Exposure model		Exposures associated with hazard of an adverse event on day *t*	Simulation scenario (S) and/or pharmacovigilance model (P)[a]	Subsets containing each pharmacovigilance model[b]
(h)	Delayed effect	Historical exposures, with the effect of exposures increasing, reaching a peak at month 3, and then decreasing over time	S	–
(i)	Decaying effect	Current exposures, with the effect of exposures decreasing over time	S	–
(j)	Decaying and delayed effect	More recent exposures and certain historical exposures	S	–
(k)	Dual effect	The timing of historical exposures: more recent exposures were assumed to be harmful, while more distant exposures were protective	S	–
(m)	Binary WCE	The weight functions were flexibly estimated over 180 days of use for cohorts simulated using models (a)–(j) and 540 days for cohorts simulated using model (k)	P	(ii) (iii) (iv) (vii) (viii)
(m.1)	Binary WCE, longer estimation period	The weight functions were flexibly estimated over 365 days of use for cohorts simulated using models (a)–(j) and 2730 days of use for cohorts simulated using model (k)	P	(vii) (viii)
(m.2)	Binary WCE, estimation period equal to duration of follow-up	The weight functions were flexibly estimated over 1095 days of use for all cohorts	P	(vii) (viii)
(n)	Continuous WCE	The weight functions were flexibly estimated over 180 days of dosages for cohorts simulated using models (a)–(j) and 540 days for cohorts simulated using model (k)	P	(ii) (iii) (v) (vii) (ix)
(n.1)	Continuous WCE, longer estimation period	The weight functions were flexibly estimated over 365 days of dosages for cohorts simulated using models (a)–(j) and 730 days of dosages for cohorts simulated using model (k)	P	(vii) (ix)
(n.2)	Continuous WCE, estimation period equal to duration of follow-up	The weight functions were flexibly estimated over 1095 days of dosages for all cohorts	P	(vii) (ix)

WCE weighted cumulative effect

[a] Exposure models used to simulate data are indicated with an 'S' (Sect. 2.1). Exposure models used to analyze the data either as single models or as part of a subset of models are indicated with a 'P' (Sect. 2.2)

[b] The pharmacovigilance models included in each pharmacovigilance subset (i)–(ix) are indicated in this column (Sect. 2.2)

aforementioned 11 exposure–risk models. For an additional scenario (l) (Table 1), in which individual drug exposure patterns and event times were matched *at random*, implying no exposure–risk association, we generated 10,000 cohorts. These cohorts were used to calibrate the stopping-rule thresholds such that the false positive rate was fixed throughout our analyses, as described in Sect. 2.3.

In the main simulations, we assumed random censoring, no exposure misclassification, complete ascertainment of AEs, and no confounding. Sensitivity analyses, outlined in Sect. 2.5, assessed the impact of these assumptions.

2.2 Prospective Pharmacovigilance Models Used to Analyze the Simulated Data

Prospective surveillance was conducted by repeating analyses at up to 20 'testing times', evenly spaced during the 5 years of follow-up to mimic a 5-year pharmacovigilance study with testing repeated every 3 months. While analyzing the data simulated from a given 'true' exposure model, we adopted a conservative assumption that the data analysts do not know how the drug exposure affects the risks and, thus, consider a wide range of potentially plausible models. Accordingly, for each simulation scenario, at each testing time, we estimated the same eight alternative prospective pharmacovigilance models, each assuming a different exposure–risk association (hereafter referred to as the 'pharmacovigilance model'). The pharmacovigilance models included six simple models used to simulate data in scenarios (a)–(f) (Table 1) and two flexible WCE models [41], in which the weight function was assumed *not* to be known and, thus, had to be estimated. Specifically, a 'binary' WCE model (m) ignored dosage and accounted only for timing and duration of past exposures, whereas a 'continuous' WCE model (n) accounted for dosage, in addition to timing and duration of exposure (Table 2). For WCE models, the weights were estimated using cubic regression B-splines with two interior knots and were constrained to zero at the end of the exposure time window, resulting in four degrees-of-freedom for the estimated exposure effect [41]. The resulting estimated effect of the exposure at any time t during follow-up was then the weighted sum of exposure indicators (for the 'binary' model) or doses (for the 'continuous' model) until time t. It should be noted that the two WCE models (m) and (n) did not assume any specific weight function and, thus, there was no one-to-one correspondence between these models and the four WCE models (g)–(k) used for data generation, each of which assumed a specific analytical weight function. Estimating an association using each of models (c), (d), (m), and (n) (Tables 1, 2) required specifying the time window over which past exposures were relevant for the current hazard. While estimating these models in the analyses, we first assumed that the time window was specified correctly, i.e., that it was identical to the window used to simulate the data for a given scenario. Then, in additional re-analyses, we also estimated alternative versions of these models, assuming a time window either longer or shorter than the true window used in data generation (Tables 1, 2).

At each testing time $T = 1, \ldots, 20$, each pharmacovigilance model used all available data from the start of surveillance until T. At each T, the hazard ratio (HR(T)) for exposure was estimated using a Cox proportional hazards model, and a likelihood ratio test statistic was used to test the null hypothesis of no association: HR(T) = 1. In Figure 4 in Electronic Supplementary Material 1, we illustrate the data structure. The corresponding p value for a given model at time T and the estimated hazard ratio were retained. Goodness-of-fit was assessed using the Akaike Information Criterion (AIC) [7, 42].

We combined evidence from pharmacovigilance models using the 'lowest p value approach' that generated a signal if the lowest model-specific p value in a pharmacovigilance subset was below the corresponding predetermined testing threshold, with the corresponding HR >1. The latter condition was imposed because typically pharmacovigilance aims at detecting only risk *increases*.

In real-life pharmacovigilance studies, depending on prior knowledge and clinical or pharmacological insights, investigators may decide to use different sets of pharmacovigilance models. Thus, we compared the results obtained with nine alternative partly overlapping pharmacovigilance subsets. The last column of Tables 1 and 2 indicates in which pharmacovigilance subsets a given model was included. Detailed descriptions of the pharmacovigilance subsets are given in Electronic Supplementary Material 3.

2.3 Calculating the Predetermined Thresholds

To ensure a comparable risk of false positive signals (α), e.g., $\alpha = 0.05$, the threshold used for 'signal detection' should correspond to the $(1-\alpha)\cdot 100^{\text{th}}$ percentile of the distribution of each respective test statistic and each pharmacovigilance subset, expected when there was no exposure-risk association (H_0). In our case, it was evident that using both (a) multiple testing times and (b) multiple pharmacovigilance models would increase false positive signals. However, the exact distribution of the test statistics was difficult to estimate analytically, as it depended not only on (a) the number of testing times and (b) the number of pharmacovigilance models in a given pharmacovigilance subset, but also on the correlations of both (c) the results of the same model at different testing times and (d) the results of different models, within a given pharmacovigilance subset, at a given T. Therefore, we relied on large-scale simulations to approximate the distribution of the test statistic [30, 41, 43] expected under H_0. These simulations were performed separately for each pharmacovigilance subset. In particular, we first simulated 10,000 cohorts, while assuming no association [model (l)], and then combined the results using a specific

pharmacovigilance subset (Sect. 2.2) applied to each simulated cohort. We then systematically increased the threshold for the test statistics and, for each such potential threshold, calculated the percentage of simulated cohorts in which H_0 was (incorrectly) rejected. This method allowed us to empirically determine a corrected threshold for each pharmacovigilance subset (Sect. 2.2), such that the corresponding observed overall false positive rate was 5%. For each pharmacovigilance subset, we observed that the distribution of *p* values and test statistics based on the 10,000 simulated cohorts remained relatively constant across all testing times (e.g., Electronic Supplementary Material 5.1.4). Therefore, while analyzing the data simulated from the models that assumed the 'true' association (Sect. 2.1), a constant threshold was used across all testing times for each pharmacovigilance subset.

2.4 Criteria to Compare the Performance of the Alternative Pharmacovigilance Subsets

Each simulated sample was re-analyzed many times using one of the pharmacovigilance subsets listed in Tables 1 and 2. Based on these results, for each of the eleven simulation scenarios (a)–(k) (Tables 1 and 2), we assessed the impact of selecting a specific pharmacovigilance subset. Specifically, we compared the time to signal detection of the 'lowest *p* value approach' across the nine pharmacovigilance subsets. Accordingly, the signal detection time for a given pharmacovigilance subset was the detection time of the best-performing model within a particular pharmacovigilance subset even if the best-performing model differed across the simulated samples. The resulting detection times were also compared against the single 'etiologically correct' pharmacovigilance model, used to generate the data for a given simulation scenario, which represented the (practically unrealistic) ideal approach. For each simulation scenario (a)–(k), times to signal detection for each pharmacovigilance subset were summarized using Kaplan–Meier-like curves [30], which show the cumulative proportion of simulated samples that generated a signal up to and including a given testing interval.

In Electronic Supplementary Material 6, we provide additional details on the results obtained with the 'lowest *p* value approach', using, as an example, pharmacovigilance subset (iii). These include descriptive statistics on the frequency distribution of the signal-generating pharmacovigilance models for each scenario, and the corresponding hazard ratios at the time of signal detection.

2.5 Sensitivity Analyses

In addition to the 'lowest *p* value approach' presented in Sect. 2.2, in Electronic Supplementary Material 5.1 we compared the timeliness of signal detection between four other approaches to pooling the *p* values of alternative pharmacovigilance models [34, 37].

Moreover, to assess the sensitivity of our results to assumptions made in conducting our main simulations, we performed four analyses. In generating data for each of the four sensitivity analyses we modified one of the following factors: (1) the variation over time in exposure patterns was decreased; (2) the incidence rate; (3) the presence of unmeasured confounding (resulting in an under-estimation of the HR for exposure) and the misclassification of exposure due to different patterns of non-adherence; and (4) the misclassification of the outcome, due to missing information on some events. Electronic Supplementary Material 5.2 presents the details of these analyses.

3 Results

The results from applying the 'lowest *p* value approach' to six different simulation scenarios are presented in Fig. 1, with additional results in Electronic Supplementary Material 4 Figures 7 and 8. Among the nine pharmacovigilance subsets (see Tables 1, 2 for composition of each pharmacovigilance subset), the three pharmacovigilance subsets (i), (iii), and (vi) performed consistently well for all ten simulation scenarios where the true model generated a harmful exposure effect for the entire exposure history, implying that drug exposure (or higher drug dose) at any time in the past (up to the current time) is always associated with risk increase (Tables 1, 2, models (a)–(j)). Each of the three pharmacovigilance subsets contained the six main parametric models (a)–(f). Pharmacovigilance subset (i) was limited to these six models, whereas pharmacovigilance subset (iii) also contained the two WCE models (m) and (n), and pharmacovigilance subset (vi) contained all ten parametric models but not the WCE models.

Among these three best performing pharmacovigilance subsets, the pharmacovigilance subset (iii) that included the two WCE models, in addition to the six parametric models, also performed well in detecting a 'dual effect' with 'crossing hazards', where the current risk *increased* with recent drug exposures but *decreased* with exposures that occurred in more distant past (more than 3 months ago) (Table 2, scenario (k)). In contrast, pharmacovigilance subsets (i) and (vi), which were limited to simple parametric models (Table 1), did not adequately detect this complex effect, as reflected by a low cumulative sensitivity (<20%) at the twentieth testing time (Electronic Supplementary Material 4, Figure 8k).

Importantly, for all simulation scenarios, the median time to detection of pharmacovigilance subset (iii)—

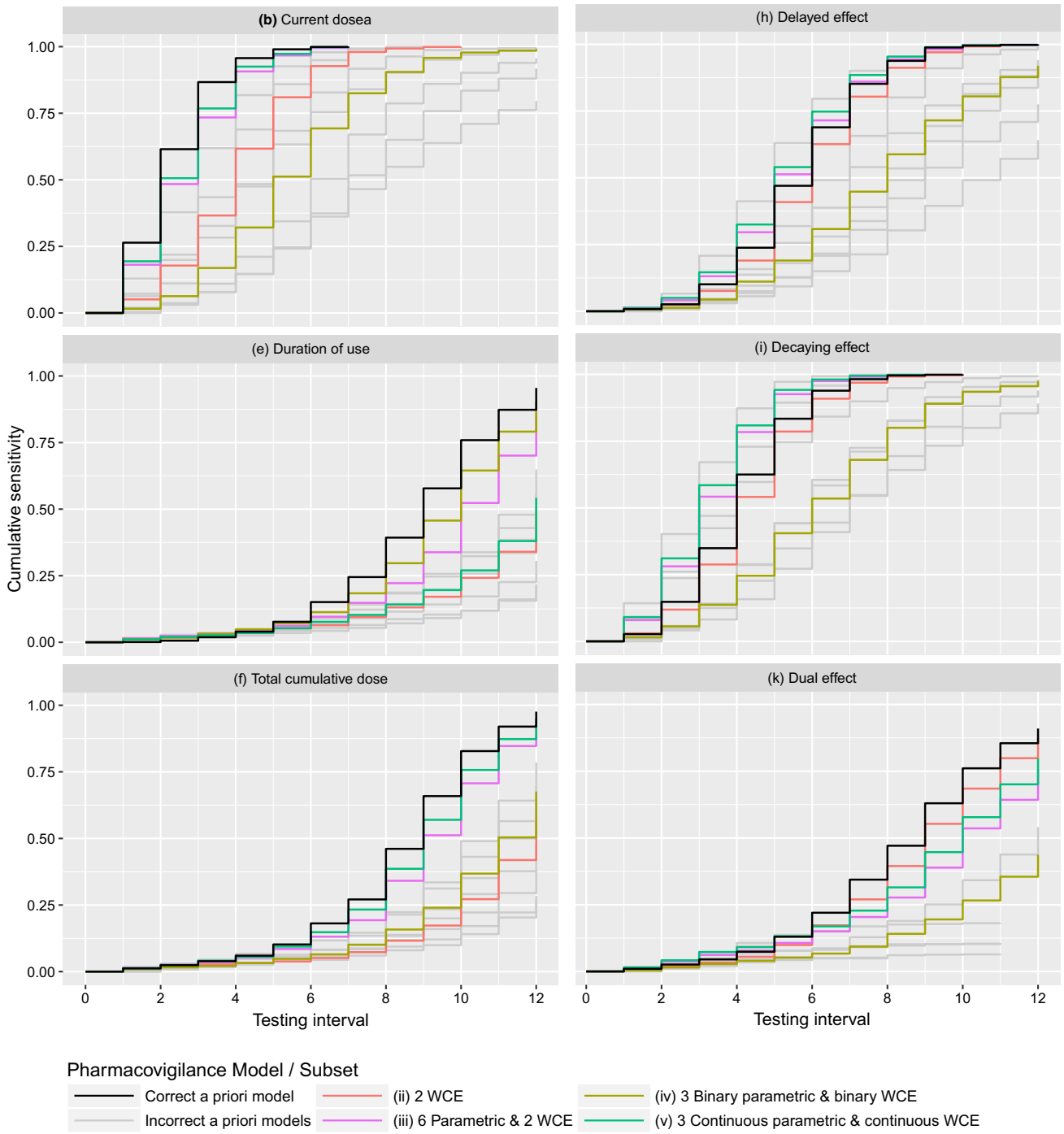

Fig. 1 Comparison of the times to signal detection generated by alternative pharmacovigilance subsets with the 'lowest p value approach'. *Each panel* corresponds to a different simulation scenario, with data generated from a different 'true' exposure model. The *horizontal axis* represents the time T (quantified by consecutive testing intervals) and the *vertical axis* shows the proportion of simulated samples in which the signal was detected by time T using a specific pharmacovigilance model/subset (with different models/subsets identified by *different colors*). For example, in **b** the signal was detected in 62% of samples by the sixth testing interval using subset (ii) (*red curve*). In *each panel*, the *solid black line* represents the pharmacovigilance model that would provide a correct interpretation of the underlying association. In **b**, **e**, and **f**, this *line* represents the pharmacovigilance model that is identical to the model used to simulate the data. In **h**, **i**, and **k**, this *line* represents a flexible weighted cumulative effect model that accounted for dosage. All other 'incorrect pharmacovigilance models' are depicted using a *solid gray line*. *WCE* weighted cumulative effect

indicated as the testing interval where the solid green line reaches the cumulative sensitivity of 0.5—was either shorter [scenario (k)] than or equivalent to the better of pharmacovigilance subsets (i) and (vi), or—on rare occasions—longer by at most only one testing interval (Electronic Supplementary Material 4, Figures 7 and 8). Although pharmacovigilance subset (iii) sometimes did not provide the timeliest detection among the nine pharmacovigilance subsets (e.g., Fig. 1e), in such (rare) cases, the difference in time to detection between the optimal pharmacovigilance subset and pharmacovigilance subset (iii) was only minor.

As expected, for scenarios simulated using a parametric model [scenarios (a)–(g)], none of the pharmacovigilance subsets performed better than the ideal (though unrealistic) approach that relied on the single 'correct pharmacovigilance model', which corresponded exactly to the true data-generating model (Fig. 1 and Electronic Supplementary Material 4, Figures 7 and 8, solid black line). For more complex models of completely harmful effects, the correct a priori model—the continuous WCE model—required more time, on average, to detect an association than the parametric model that most closely resembled the underlying complex association. The additional degrees of freedom (*df*) associated with WCE models results in an additional penalty for the model's complexity, which decreases the power of the test of association [7] such that the 1 *df* simple, parametric model that closely resembles the complex 'true' association was a more timely indicator of the presence of an association (Fig. 1h, i). For example, for the current dose scenario (b), the timeliest signal detection was obtained using the current dose model (b); for the delayed effect scenario (h), the timeliest signal detection was obtained using the model where the relevant exposure was the cumulative dose in the past 30 days [model (d)].

Our results in Fig. 1 suggest that, if the true model is *not* known, an association can be detected in a more timely manner by fitting a subset (iii) of pharmacovigilance models. The set of models included alternative parametric and flexible models with different assumptions about the underlying exposure–risk association.

In our sensitivity analyses, we compared the times to detection between the 'lowest p value approach' and four alternative approaches for combining evidence from multiple pharmacovigilance models for each of the 11 simulation scenarios and each of the nine model pharmacovigilance subsets. Figures 9 and 10 in Electronic Supplementary Material 5.1.3 illustrate these results for pharmacovigilance subset (iii). Our results suggest that more sophisticated approaches to pooling evidence from multiple pharmacovigilance models do not significantly improve the timeliness of signal detection compared with the simple 'lowest p value approach'. Further sensitivity analyses confirmed findings of our main simulations: among the alternative approaches, the 'lowest p value approach' consistently offered the fastest signal detection and highest sensitivity of signal detection by the last testing interval. Our analyses suggested that the original thresholds may be conservative when (1) exposure patterns were stable relative to the cohorts used in the original analysis; (2) the incidence rate was higher than in the original simulations; or (3) there was a combination of unmeasured confounding, resulting in an under-estimation of the true HR for the exposure, and misclassification of exposure due to different patterns of non-adherence. In contrast, when information on some events was missing, false positive rates tended to be slightly higher. Among the four modifications of data structure, the most significant detection delays were observed in the presence of unmeasured confounding resulting in an under-estimation of the true HR for the exposure, combined with misclassification of exposure due to different patterns of non-adherence. Full results of these sensitivity analyses are presented in Electronic Supplementary Material 5.2.

4 Discussion

In this article, we used simulation to investigate the potential benefits of combining results of multiple exposure–risk models. We hypothesized that simultaneously using the results of several exposure models could offer more timely detection, relative to an arbitrarily selected model selected a priori. Accordingly, we combined evidence from multiple exposure models applied to the same simulated data, and compared the resulting times to signal detection. To conduct this analysis, we simulated various plausible patterns of association between time-varying drug use and the hazard of an AE. We then analyzed the simulated data assuming that, as in most real-life prospective surveillance studies, the 'true' data-generating exposure–risk model is unknown. We compared the resulting times to signal detection under a range of simulated, clinically plausible associations between a time-varying drug exposure and the hazard of an AE. We thereby demonstrated the impact of methodological choices in modeling drug exposure on the timeliness of signal detection in pharmacovigilance [30].

Consistent with our expectations, the 'lowest p value approach', which considered evidence from multiple models, detected signals of association at a time comparable to an unrealistic 'ideal case' scenario, where only the single 'correct' pharmacovigilance model, corresponding to the 'true' model used to simulate the data, was employed. Conversely, combining results of alternative

models improved both the timeliness and the sensitivity of AE detection at a fixed specificity as compared to standard current pharmacovigilance practice, where a single model is selected arbitrarily a priori and used to assess the putative association [44–49]. The additional computational burden, required to estimate several alternative models and combine their results, appears to be justifiable as it compensates for uncertainty about the true exposure–risk model. The 'lowest p value approach' generates a signal when the p value of at least one of the pharmacovigilance models considered in the analyses falls below a certain predetermined corrected threshold. The p value threshold was derived through simulations [43, 50] and corrected for false positive rate inflation due to both testing repeated over time and the use of multiple exposure models to ensure an overall false positive rate of about 5%.

Overall, our results indicate that, in situations when the 'true' model, which specifies how past and current drug exposures affect the risk of a particular AE, is *not* known, signal detection is more timely when using a pharmacovigilance subset based on distinct yet plausible assumptions than when using any single model. As shown in Fig. 1, the best overall results were obtained with the pharmacovigilance subset (iii) that included alternative plausible models that assumed different exposure–risk associations and flexible WCE models that made no assumptions regarding the form/shape of the association. The results of our sensitivity analyses provide reasonable assurances that, in most situations, the actual false positive rate is likely to be equal to or lower than the prespecified nominal rate for which the threshold was set.

Our study has several limitations. First, our simulation scenarios considered only a selection of the models that may represent the true relationships between different drug exposures and various AEs, although we sought to include a representative set of clinically plausible associations of interest. Second, our simulations only allowed us to compare the *relative* performance of alternative approaches, in terms of faster or slower times to signal detection. They do not allow us to predict the actual magnitude of absolute time reduction in specific real-life situations, as the time to detection will depend on such factors as the amount of available data (e.g., the sample size, prevalence, and time-varying patterns of use of the drug of interest and the outcome rate, which will determine the number of events), frequency of testing times, variability in drug use, and confounders adjusted for in the analysis. Third, we somewhat arbitrarily selected the 'target' false positive rate of 5%, chose our threshold for signal detection to be constant over time, and pre-set the number of testing intervals. However, we showed that the thresholds we generated are robust to changes in exposure patterns, the incidence rate, and outcome misclassification, suggesting that these thresholds may produce similar false positive rates in future studies of different drug exposures and/or different AEs. Future work is needed, however, to characterize more precisely the range of conditions under which the thresholds generated in the present study remain valid. In situations where a different set of models may be required, and/or where the sample size and study design diverge substantially from those considered in our simulations, it might be preferable to conduct 'customized' simulations and derive new thresholds based on the results. In such situations the 'generic' methods described in Sect. 2.3 of our article may be useful.

Importantly, methods to pool p values that we considered in this paper have traditionally been used to combine information from multiple genes [36, 51–53] or multiple trials focusing on the same intervention [37]. In such previous applications, each p value was derived from a different information-generating source (e.g., different genetic markers, genes, DNA segments, or independently conducted trials). This is not the case when alternative pharmacovigilance models are used to assess a single relationship using exactly the same data, which implies that the amount of additional information gained by pooling results of different analyses may be lower. However, we believe it is important to provide solid empirical evidence for assessing the relative advantages of different approaches to pooling results of alternative pharmacovigilance models. The fact that a simple, easy-to-implement 'lowest p value approach' performed the best across a range of simulation scenarios adds to the practical relevance of our findings.

5 Conclusions

Prospective pharmacovigilance aims to detect safety concerns related to medical products in a timely manner. Taken together, our results suggest that, in typical applications where the 'true' exposure–risk model is unknown a priori, an effect can be detected in a more timely manner using a carefully selected set of pharmacovigilance models and a simple approach to combining their results. Specifically, the approach we recommend requires first fitting the selected set of alternative pharmacovigilance models and then relying on the model that generates the earliest signal, while using a simulation-corrected p value cut-off that ensures adequate control of the overall false positive rate. Our results suggest that an optimal pharmacovigilance subset should include (a) both dose-dependent models and models that ignore the dosage; (b) models of varying durations of effect; and (c) both parametric and flexible WCE models [41, 54–56]. Implementation of this approach in real-life pharmacovigilance will provide additional insights into its potential practical advantages and limitations, and will likely stimulate further method developments.

Acknowledgements The authors thank two anonymous reviewers whose thoughtful reviews improved this work.

Compliance with Ethical Standards

Funding This research was funded by the Canadian Institutes for Health Research (CIHR) Grant MOP-81275.

Conflict of interest Prior to May 2014, Rolina van Gaalen 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. Michal Abrahamowicz and David Buckeridge have no conflicts of interest that are directly relevant to the content of this study.

References

1. Harpaz R, DuMouchel W, LePendu P, Bauer-Mehren A, Ryan P, Shah NH. Performance of pharmacovigilance signal-detection algorithms for the FDA adverse event reporting system. Clin Pharmacol Ther. 2013;93:539–46.
2. Franklin JM, Schneeweiss S, Polinski JM, Rassen JA. Plasmode simulation for the evaluation of pharmacoepidemiologic methods in complex healthcare databases. Comput Stat Data Anal. 2014;72:219–26.
3. Kesselheim AS, Gagne JJ. Strategies for post marketing surveillance of drugs for rare diseases. Clin Pharmacol Ther. 2014;95:265–8.
4. Trifiro G, Coloma PM, Rijnbeek PR, Romio S, Mosseveld B, Weibel D, et al. Combining multiple healthcare databases for postmarketing drug and vaccine safety surveillance: why and how? J Intern Med. 2014;275:551–61.
5. Bartlett G, Abrahamowicz M, Grad R, Sylvestre M-P, Tamblyn R. Association between risk factors for injurious falls and new benzodiazepine prescribing in elderly persons. BMC Fam Pract. 2009;10:1.
6. Christensen L, Sasané R, Hodgkins P, Harley C, Tetali S. Pharmacological treatment patterns among patients with attention deficit/hyperactivity disorder: retrospective claims-based analysis of a managed care population. Curr Med Res Opin. 2010;26:977–89.
7. Abrahamowicz M, Beauchamp M-E, Sylvestre M-P. Comparison of alternative models for linking drug exposure with adverse effects. Stat Med. 2012;31:1014–30.
8. Nelson JC, Cook AJ, Onchee Y, Zhao S, Jackson LA, Psaty BM. Methods for observational post-licensure medical product safety surveillance. Stat Methods Med Res. 2015;24:177–93.
9. McMahon AD, Evans JMM, McGilchrist MM, McDevitt DG, MacDonald TM. Drug exposure risk windows and unexposed comparator groups for cohort studies in pharmacoepidemiology. Pharmacoepidemiol Drug Saf. 1998;7:275–80.
10. Csajka C, Verotta D. Pharmacokinetic–pharmacodynamic modelling: history and perspectives. J Pharmacokinet Pharmacodyn. 2006;33:227–79.
11. Karimi G, Star K, Noren NG, Hagg S. The impact of duration of treatment of reported time-to-onset in spontaneous reporting systems for pharmacovigilance. PLOS One. 2013;8:e68938.
12. Kiri VA. A pathway to improved prospective observational post-authorization safety studies. Drug Saf. 2012;35:711–24.
13. Fireman B, Toh S, Butler MG, Go AS, Joffe HV, Graham DJ, 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;21:282–90.
14. Harpaz R, DuMouchel W, Shah NH, Madigan D, Ryan P, Friedman C. Novel data-mining methodologies for adverse drug event discovery and analysis. Clin Pharmacol Ther. 2012;91:1010–21.
15. Motsko SP, Rascati KL, Busti AJ, Wilson JP, Barner JC, Lawson KA, et al. Temporal relationship between use of NSAIDs, including selective COX-2 inhibitors, and cardiovascular risk. Drug Saf. 2006;29:621–32.
16. Helin-Salmivaara A, Virtanen A, Vesalainen R, Grönroos JM, Klaukka T, Idänpään-Heikkilä JE, et al. NSAID use and the risk of hospitalization for first myocardial infarction in the general population: a nationwide case–control study from Finland. Eur Heart J. 2006;27:1657–63.
17. Ray WA, Stein CM, Hall K, Daugherty JR, Griffin MR. Non-steroidal anti-inflammatory drugs and risk of serious coronary heart disease: an observational cohort study. Lancet. 2002;359:118–23.
18. Ray WA, Varas-Lorenzo C, Chung CP, Castellsague J, Murray KT, Stein CM, et al. Cardiovascular risks of nonsteroidal anti-inflammatory drugs in patients after hospitalization for serious coronary heart disease. Circ Cardiovasc Qual Outcomes. 2009;2:155–63.
19. García Rodríguez LA, Varas-Lorenzo C, Maguire A, González-Pérez A. Nonsteroidal antiinflammatory drugs and the risk of myocardial infarction in the general population. Circulation. 2004;109:3000–6.
20. García Rodríguez LA, Tacconelli S, Patrignani P. Role of dose potency in the prediction of risk of myocardial infarction associated with nonsteroidal anti-inflammatory drugs in the general population. J Am Coll Cardiol. 2008;52:1628–36.
21. Tournier M, Bégaud B, Cougnard A, Auleley G-R, Deligne J, Blum-Boisgard C, et al. Influence of the drug exposure definition on the assessment of the antipsychotic metabolic impact in patients initially treated with mood-stabilizers. Br J Clin Pharmacol. 2012;74:189–96.
22. Østbye T, Curtis LH, Masselink LE, Hutchison S, Wright A, Dans PE, et al. Atypical antipsychotic drugs and diabetes mellitus in a large outpatient population: a retrospective cohort study. Pharmacoepidemiol Drug Saf. 2005;14:407–15.
23. Guo JJ, Keck PE, Corey-Lisle PK, Li H, Jiang D, Jang R, et al. Risk of diabetes mellitus associated with atypical antipsychotic use among Medicaid patients with bipolar disorder: a nested case–control study. Pharmacotherapy. 2007;27:27–35.
24. Cook AJ, Tiwari RC, D'Wellman R, Heckbert SR, Li L, Heagerty PJ, 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.
25. Brown JS, Kulldorff M, Chan KA, Davis RL, Graham D, Pettus PT, et al. Early detection of adverse drug events within population-based health networks: application of sequential testing methods. Pharmacoepidemiol Drug Saf. 2007;16:1275–84.
26. Gagne JJ, Rassen JA, Walker AM, Glynn RJ, Schneeweiss S. Active safety monitoring of new medical products using electronic healthcare data: selecting alerting rules. Epidemiology. 2012;23:238–46.
27. Perucca P, Gilliam FG. Adverse effects of antiepileptic drugs. Lancet Neurol. 2012;11:792–802.
28. Stricker BHC, Stijnen T. Analysis of individual drug use as a time-varying determinant of exposure in prospective population-based cohort studies. Eur J Epidemiol. 2010;25:245–51.
29. Zaccara G, Franciotta D, Perucca E. Idiosyncratic adverse reactions to antiepileptic drugs. Epilepsia. 2007;48:1223–44.
30. Van Gaalen RD, Abrahamowicz M, Buckeridge D. The impact of exposure model misspecification on the timeliness of signal

detection in prospective pharmacovigilance. Pharmacoepidemiol Drug Saf. 2015;24:456–67.
31. Volinsky CT, Madigan D, Raftery AE, Kronmal RA. Bayesian model averaging in proportional hazard models: assessing the risk of a stroke. Appl Stat. 1997;46:433–48.
32. Hoeting JA, Madigan D, Raftery AE, Volinsky CT. Bayesian model averaging: a tutorial. Stat Sci. 1999;14:382–401.
33. Claeskens G, Hjort NL. Model selection and model averaging. Cambridge: Cambridge University Press; 2008.
34. Chen Z. Is the weighted z-test the best method for combining probabilities from independent tests? J Evol Biol. 2011;24:926–30.
35. Zaykin DV, Zhivotovsky LA, Westfall PH, Weir BS. Truncated product method for combining p-values. Genet Epidemiol. 2002;22:170–85.
36. Delongchamp R, Lee T, Velasco C. A method for computing the overall statistical significance of a treatment effect among a group of genes. BMC Bioinform. 2006;7(Suppl. 2):S11.
37. Zaykin DV. Optimally weighted Z-test is a powerful method for combining probabilities in meta analysis. J Evol Biol. 2011;24:1836–41.
38. Sylvestre M-P, Evans T, MacKenzie T, Abrahamowicz M. PermAlgo: permutational algorithm to generate event times conditional on a covariate matrix including time-dependent covariates. 2010; R package version 1.1..http://cran-mirror.cs.uu.nl/. Accessed 15 June 2017.
39. Sylvestre M-P, Abrahamowicz M. Comparison of algorithms to generate event times conditional on time-dependent covariates. Stat Med. 2008;27:2618–34.
40. 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.
41. Sylvestre M-P, Abrahamowicz M. Flexible modeling of the cumulative effects of time-dependent exposures on the hazard. Stat Med. 2009;28:3437–53.
42. Akaike H. A new look at the statistical model identification. IEEE Trans Autom Control. 1974;19:716–23.
43. Mahmud M, Abrahamowicz M, Leffondré K, Chaubey YP. Selecting the optimal transformation of a continuous covariate in Cox's regression: implications for hypothesis testing. Commun Stat Simul Comput. 2006;35:27–45.
44. Saag KG, Koehnke R, Caldwell JR, Brasington R, Burmeister LF, Zimmerman B, et al. Low dose long-term corticosteroid therapy in rheumatoid arthritis: an analysis of serious adverse events. Am J Med. 1994;96:115–23.
45. Franklin J, Lunt M, Symmons D, Silman A. Risk and predictors of infection leading to hospitalisation in a large primary-care-derived cohort of patients with inflammatory polyarthritis. Ann Rheum Dis. 2007;66:308–12.
46. Lacaille D, Guh DP, Abrahamowicz M, Anis AH, Esdaile JM. Use of nonbiologic disease-modifying antirheumatic drugs and risk of infection in patients with rheumatoid arthritis. Arthritis Rheum. 2008;59:1074–81.
47. Schneeweiss S, Setoguchi S, Weinblatt ME, Katz JN, Avorn J, Sax PE, et al. Anti-tumor necrosis factor alpha therapy and the risk of serious bacterial infections in elderly patients with rheumatoid arthritis. Arthritis Rheum. 2007;56:1754–64.
48. Smitten AL, Choi HK, Hochberg MC, Suissa S, Simon TA, Testa MA, et al. The risk of hospitalized infection in patients with rheumatoid arthritis. J Rheumatol. 2008;35:387–93.
49. Bernatsky S, Hudson M, Suissa S. Anti-rheumatic drug use and risk of serious infections in rheumatoid arthritis. Rheumatology. 2007;46:1157–60.
50. 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 Simul Comput. 2009;38:1115–35.
51. Dudbridge F, Koeleman BPC. Efficient computation of significance levels for multiple associations in large studies of correlated data, including genome-wide association studies. Am J Hum Genet. 2004;75:424–35.
52. Conneely KN, Boehnke M. So many correlated tests, so little time! Rapid adjustment of p values for multiple correlated tests. Am J Hum Genet. 2007;81:1158–68.
53. Dai H, Leeder JS, Cui Y. A modified generalized Fisher method for combining probabilities from dependent tests. Front Genet. 2014;5:32.
54. Dixon WG, Abrahamowicz M, Beauchamp M-E, Ray DW, Bernatsky S, Suissa S, 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;71:1128–33.
55. Young J, Xiao Y, Moodie EEM, Abrahamowicz M, Klein M, Bernasconi E, et al. Effect of cumulating exposure to abacavir on the risk of cardiovascular disease events in patients from the Swiss HIV Cohort Study. J Acquir Immune Defic Syndr. 2015;69:413–21.
56. Movahedi M, Beauchamp ME, Abrahamowicz M, Ray DW, Michaud K, Pedro S, et al. Risk of incident diabetes associated with dose and duration of oral glucocorticoid therapy in patients with rheumatoid arthritis. Arthritis Rheumatol. 2016;68:1089–98.