STATISTICS IN MEDICINE
Statist. Med. 2000; **00**:1–18 *Prepared using* simauth.cls *[Version: 2002/09/18 v1.11]*

Extending the methods used to screen the WHO drug safety database towards analysis of complex associations and improved accuracy for rare events

G. Niklas Norén[1,2,*], Andrew Bate[1], Roland Orre[3] and I. Ralph Edwards[1]

[1] *The WHO Collaborating Centre for International Drug Monitoring in Uppsala, Sweden*
[2] *Mathematical statistics, Stockholm University, Stockholm, Sweden*
[3] *NeuroLogic Sweden AB, Stockholm, Sweden*

SUMMARY

Post-marketing drug safety data sets are often massive, and entail problems with heterogeneity and selection bias. Nevertheless, quantitative methods have proven a very useful aid to help clinical experts in screening for previously unknown associations in these data sets. The WHO international drug safety database is the world's largest data set of its kind with over 3 million reports on suspected adverse drug reaction incidents. Since 1998, an exploratory data analysis method has been in routine use to screen for quantitative associations in this data set. This method was originally based on large sample approximations and limited to pairwise associations, but in this article we propose more accurate credibility interval estimates and extend the method to allow for the analysis of more complex quantitative associations. The accuracy of the proposed credibility intervals is evaluated through comparison to precise Monte Carlo simulations. In addition, we propose a Mantel-Haenszel type adjustment to control for suspected confounders. Copyright © 2000 John Wiley & Sons, Ltd.

1. Introduction

Despite great efforts in investigating drug safety before new substances are introduced on the market, some adverse drug reactions (ADR) are not detected until after drug launch. This applies in particular to reactions that have low incidence, occur primarily in groups that tend to be excluded from clinical trials (such as pregnant women or young children), are due to drug interactions or have long times to onset [1]. Screening of spontaneous reports is one of several tools for post-marketing drug safety surveillance [2, 3], and remains the main method for generating hypotheses related to previously unknown adverse drug reactions [4, 5]. In this context, international initiatives have the advantage of accumulating information from all over the world, something which increases the potential for early detection of drug safety problems [6]. At the same time, the massive data sets involved require quantitative methods for efficient knowledge discovery.

The WHO Collaborating Centre for International Drug Monitoring in Uppsala, Sweden (also known as the Uppsala Monitoring Centre or the UMC) holds the world's largest database of

*Correspondence to: Niklas Norén, niklas.noren@who-umc.org

spontaneous reports on suspected adverse drug reaction incidents. The first reports in this data set date back to 1967, and as of November 2004, 75 countries from around the world forward their ADR reports to the UMC. The database currently consists of over 3 million reports, with more than 14,000 distinct drug substances and almost 2,000 distinct ADR terms. Out of the over 20 million possible combinations of one of these drug substances with one of these ADR terms, around 600,000 pairs occur together on at least one report in this data set.

In recent years, several methods for quantitative analysis of spontaneous reporting data sets have been proposed — some Bayesian [7, 9] and others non-Bayesian [11, 12]. Unlike earlier approaches [13, 14], the lack of readily accessible and reliable international usage data has lead these methods to focus on associations within the data sets rather than on proper rates of incidence. Instead of external data, the whole database of reported ADR incidents is thus the reference against which each possible association is compared. Despite the biases in this reference population and despite biases in reporting behaviour and problems with data quality (*e.g.* the highly variable amount of information available on different reports and the phenomenon of duplicate reports), these quantitative methods have proven a useful aid in highlighting drug-ADR combinations for clinical review [15, 16]. One advantage with the study of associations within the data set is that some biases (such as the relative over-reporting of new drug substances) is automatically compensated for [12].

Let dependency derivation denote the screening for quantitative associations between events in a large data set. An obvious difficulty with exploratory data analysis ventures is the multiple comparisons problem: given the large number of possible associations that are evaluated simultaneously, it is hard to attribute a degree of significance to the findings. In addition, for drug safety data sets, even if significant quantitative associations between two events can be identified, they could potentially be driven by confounding variables or reporting biases. As a consequence, we use dependency derivation as a means not to draw final conclusions about possible associations between events in the data set but to generate hypotheses. Any suspicion raised through dependency derivation needs to be further evaluated and tested in some follow-up procedure. In the routine screening of the WHO database for potential drug safety problems, this follow-up procedure consists of clinical evaluation of the individual case reports for each highlighted association, by an international panel of drug safety experts [6].

In screening the database, we are interested in both the estimated strengths of association and the support in data; strong associations are more likely to be indicative of important problems, but with very little support in data even strong quantitative associations are likely to be spurious. The problem with the straightforward use of a test for association is that it may tend to highlight weak associations with large data support [9]. On the other hand, raw strength of association estimates are sensitive to random variation and thus vulnerable to spurious associations. In fact, for these large and sparse data sets, even the use of classical confidence intervals around traditional strength of association measures are insufficient to compensate for a limited data support. As an illustration, consider a drug substance x for which in the first quarter after marketing there were only two case reports in the WHO database. Even for a common ADR term, with say 100,000 reports in total, a single observed report together with this drug by far exceeds the expected number (0.07) and the 95% confidence interval for the log-odds ratio ([0.60 6.14]) excludes 0 and thus indicates a quantitative association. Consequently, if log-odds ratios with confidence intervals were used as the screening criterion, single reports on a particular ADR would suffice to highlight new drugs for clinical review.

Bayesian dependency derivation methods can be used to provide a reasonable balance

between strength of association and support in data [7, 9]. Bayesian inference is sometimes criticised on account of the explicit incorporation of prior assumptions in the analysis, but in dependency derivation this is the greatest advantage over classical inference. The conservative prior distribution (based on the á priori assumption of mutual independence between any two events) moderates the strength of association estimates toward the baseline assumption of no association (especially at low counts) and thereby reduces the risk of highlighting spurious associations.

Since 1998, a Bayesian dependency derivation method has been in routine use to rank quantitative associations between drug substances and ADR terms in the WHO database [7, 8]. The use of this method to highlight drug-ADR combinations for clinical review has been thoroughly tested [15] and integrated into the overall signal detection strategy at the UMC [16, 17, 18, 19]. Several associations first highlighted with this approach have been published in the medical literature [20, 21]. The algorithmic framework used is referred to as the Bayesian Confidence Propagation Neural Network (BCPNN). The BCPNN is a statistical neural network where the nodes correspond to different events and the weights between nodes are proportional to the strength of association between different events. The BCPNN can be used for complex tasks such as classification and unsupervised pattern recognition [22, 23, 24, 25, 26], but for the purpose of dependency derivation, only the weights between nodes of the network (referred to as Information Components or IC values) are of interest. These can be estimated directly from data, so for transparency we shall refer to the use of the BCPNN for Bayesian dependency derivation as IC analysis throughout this article.

In this article, we propose more accurate credibility interval estimates for the prior/posterior distribution of the IC that do not rely on large sample theory. We argue in favour of using the mode as the central IC estimate and show how it can be accurately estimated. In addition, we propose a generalisation of the IC to higher order associations in order to screen for ADR risk factors. We also introduce a Mantel-Haenszel type of adjustment for the IC in order to control for potential confounders in heterogeneous data sets.

2. IC analysis for pairwise dependency derivation

Denote by IC_{xy} the Information Component between events x and y for variables X and Y respectively. The IC is defined as the base 2 logarithm of an observed-to-expected ratio for the joint probability of the two events, where the expected value is calculated under the assumption of mutual independence [7, 8]:

$$IC_{xy} = \log_2 \frac{P(x,y)}{P(x)P(y)} \tag{1}$$

A positive IC value indicates that the two events co-occur more frequently than expected under the assumption of independence, and a negative IC value indicates that they co-occur more rarely. The IC is a function of the unknown probabilities $P(x,y)$, $P(x)$ and $P(y)$, and Bayesian inference is used to estimate the IC value. For convenience, a Dirichlet prior distribution (that is conjugate to the multinomial distribution of data) is used for the probability parameters, since this makes closed form expressions for the posterior distributions of $P(x,y)$, $P(x)$ and $P(y)$ readily available. No such closed form expression is known for the posterior distribution of the IC itself, but in recent work, the use of Monte Carlo simulation based on the closed

form expressions for $P(x, y)$, $P(x)$ and $P(y)$, has been effective in learning more about the shape of the posterior IC distribution [27]. We use this approach to evaluate the accuracy of the approximations proposed in this article.

To simplify the annotation: with respect to the presence or absence of two events x and y, denote by p_{11}, $p_{1\cdot}$ and $p_{\cdot 1}$ the probability parameters for $P(x, y)$, $P(x)$ and $P(y)$, respectively. Similarly, denote by n_{11}, $n_{1\cdot}$ and $n_{\cdot 1}$ the corresponding numbers of observations in the data set. In addition, denote by n_{10} the number of cases where $X = x$ but $Y \neq y$, by n_{01} the number of cases where $X \neq x$ but $Y = y$, and by n_{00} the number of cases where both $X \neq x$ and $Y \neq y$. Denote by p_{10}, p_{01} and p_{00} the corresponding probabilities.

In our data model, the observed counts n_{11}, n_{10}, n_{01} and n_{00} are assumed to follow a $Mn(p_{11}, p_{10}, p_{01}, p_{00}, n_{\cdot\cdot})$ distribution. With a $Di(\alpha_{11}, \alpha_{10}, \alpha_{01}, \alpha_{00})$ prior distribution for p_{11}, p_{10}, p_{01} and p_{00}, it is a standard result from Bayesian statistics that the corresponding posterior distribution is $Di(\gamma_{11}, \gamma_{10}, \gamma_{01}, \gamma_{00})$, where $\gamma_{ij} = \alpha_{ij} + n_{ij}$ (in an abstract sense, the hyper parameters α_{ij} can be thought of as assumed prior observations) [28].

In this model, the marginal distributions of p_{11}, p_{10}, p_{01} and p_{00} are beta. The same is true for $p_{1\cdot} = p_{11} + p_{10}$ and $p_{\cdot 1} = p_{11} + p_{01}$. Specifically:

$$\begin{aligned} p_{11} &\sim Be(\gamma_{11}, \gamma_{10} + \gamma_{01} + \gamma_{00}) \\ p_{1\cdot} &\sim Be(\gamma_{11} + \gamma_{10}, \gamma_{01} + \gamma_{00}) \\ p_{\cdot 1} &\sim Be(\gamma_{11} + \gamma_{01}, \gamma_{10} + \gamma_{00}) \end{aligned} \tag{2}$$

However, since p_{11}, $p_{1\cdot}$ and $p_{\cdot 1}$ are not independent ($p_{1\cdot} = p_{11} + p_{10}$ and $p_{\cdot 1} = p_{11} + p_{01}$), it will sometimes be a coarse approximation to consider the marginal distributions separately as has been done earlier [7, 8], and we will in this article base our analyses on the full Dirichlet distribution.

Some general problems with observed-to-expected ratios should be kept in mind. Observed-to-expected ratios are relevant strength of association measures primarily for events with low expected frequencies where there is virtually no upper limit to the observed-to-expected ratios. In contrast, if the overall frequency of a certain ADR term is as high as, for example, 0.5, the observed-to-expected ratio for its association with a given drug substance can never exceed 2 – even if that ADR term occurs on every report for that drug substance. As a consequence, comparisons between IC values can potentially be misleading if the expected frequencies vary significantly in magnitude. Another problem with observed-to-expected ratios is that there may be a spill-over effect from a large observed number of reports for an event pair to the expected number of reports for that event pair. Specifically, if the drug substance under study is very common and there are unexpectedly many reports on this drug substance with a particular ADR, this may influence the overall prevalence of that ADR term so much that the strength of association is underestimated by the observed-to-expected ratio. These two problems rarely affect pairwise IC analysis between drug substances and ADR terms in standard drug safety data sets much, but may be important in the analysis of other types of events or of smaller data sets. To minimise the risk for misleading results, it may in some situations be sensible to accompany the estimated IC values for highlighted associations with standard log-odds ratios.

2.1. The moderating prior distribution

The aim of IC analysis is to generate useful leads with respect to quantitative associations in a data set. As previously discussed, it is in this context crucial to avoid the highlighting

of an abundance of associations with weak support in data, but at the same time focusing on estimated strength of association. With respect to this issue, Bayesian dependency derivation based on a conservative prior distribution has proven instrumental in moderating the estimated strengths of association when data is scarce [7, 9]. The Bayesian moderation in combination with the use of credibility intervals provides an efficient, pragmatic compromise between methods based on statistical significance only (that may be sensitive to weak associations with large data support) and methods based on raw observed-to-expected ratios (that tend to highlight associations with very little data support). Since the impact of the prior distribution diminishes as data accumulates, for combinations with large support there is little difference between Bayesian and classical estimates.

To ascertain moderation of the posterior distribution toward the baseline assumption of independence ($IC = 0$) for all possible associations in all possible data sets, assume a $Di(\alpha_{11}, \alpha_{10}, \alpha_{01}, \alpha_{00})$ prior distribution for p_{11}, p_{10}, p_{01} and p_{00} where:

$$\begin{aligned}
\alpha_{11} &= q_{1\cdot} q_{\cdot 1} \cdot \alpha_{\cdot\cdot} \\
\alpha_{10} &= q_{1\cdot} q_{\cdot 0} \cdot \alpha_{\cdot\cdot} \\
\alpha_{01} &= q_{0\cdot} q_{\cdot 1} \cdot \alpha_{\cdot\cdot} \\
\alpha_{00} &= q_{0\cdot} q_{\cdot 0} \cdot \alpha_{\cdot\cdot}
\end{aligned} \tag{3}$$

and:

$$\alpha_{\cdot\cdot} = \frac{0.5}{q_{1\cdot} q_{\cdot 1}} \tag{4}$$

and:

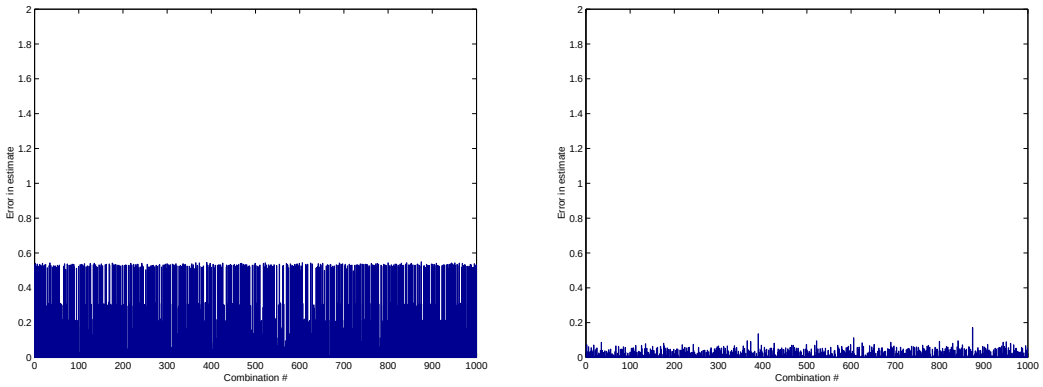
$$\begin{aligned}
q_{1\cdot} &= \frac{n_{1\cdot} + 1/2}{n_{\cdot\cdot} + 1} \\
q_{0\cdot} &= \frac{n_{0\cdot} + 1/2}{n_{\cdot\cdot} + 1} \\
q_{\cdot 1} &= \frac{n_{\cdot 1} + 1/2}{n_{\cdot\cdot} + 1} \\
q_{\cdot 0} &= \frac{n_{\cdot 0} + 1/2}{n_{\cdot\cdot} + 1}
\end{aligned} \tag{5}$$

This prior distribution incorporates the independence assumption by setting the hyper parameters proportional to the products of the corresponding marginal probabilities (in fact to posterior mean estimates for the marginal probabilities based on $Be(1/2, 1/2)$ hyper priors). The benefit of this is that the IC_{map} always lies between 0 and the raw observed-to-expected log-ratio and that:

$$\begin{aligned}
\lim_{n_{1\cdot}, n_{\cdot 1} \to 0} IC_{map} &\approx 0 \\
\lim_{n_{\cdot\cdot} \to 0} IC_{map} &= 0
\end{aligned} \tag{6}$$

which is important for computational stability.

In the abstract sense mentioned above, the moderating prior distribution is equivalent to an assumed extra batch of data where the two events under study are independent, co-occur 0.5



a. Posterior mean estimate

b. Maximum à posteriori estimate

Figure 1. Errors in posterior mean and maximum à posteriori estimates for the IC distribution. This figure shows deviations from the Monte Carlo simulated values of estimates based on (7) for the mean and the mode, respectively. 1,000 randomly selected drug-ADR pairs in the WHO database were used and each Monte Carlo simulation was based on 50,000 draws.

times and where the marginal probabilities for the two events are approximately the same as in the real data set. With this approach, the prior sample size $\alpha_{..}$ may vary, but α_{11} always equals 0.5 and since it is primarily α_{11} that determines the shape of the IC distribution [27], the shape of the prior distribution will be approximately the same for all associations under study.

This prior distribution is based on the same principles as the prior previously used in IC analysis (see [7] or [8]). The major differences are that the new prior is based on the joint Dirichlet distribution for the model parameters (instead of independent beta distributions) and that the prior sample size has been halved. The reduction in prior sample size yields a more diffuse prior distribution that better reflects our initial uncertainty about the IC values in the WHO database. For data sets with different characteristics (size, sparsity, heterogeneity) than the WHO database, the factor 0.5 in the expression for $\alpha_{..}$ should be adjusted. It may, for example, be sensible to reduce this factor (and thus the moderating effect of the prior distribution) for smaller data sets.

2.2. Central IC estimates

Arbitrarily accurate estimates for the posterior mean (p.m.e.) of the IC distribution are available [29]. However, as the IC distribution is generally unimodal, maximum á posteriori (m.a.p.) estimates may be used for central estimates instead. The main advantage of the m.a.p. estimate is that it is well suited for use in stratified IC analysis (see Section 2.4) and that it has the intuitive property of being equal 0 when the estimated joint probability equals the product of the estimated marginal probabilities. In addition, the concept of a most likely value for an unknown parameter is perhaps more natural than that of an expected value, and this is an important aspect in the drug safety application, where the results must be interpretable for non-statisticians.

	$n_{11}=1$	$n_{11}=2$	$n_{11}=3$	$n_{11}=4$	$n_{11}=5$
IC_{pme}	0.53	0.30	0.21	0.16	0.13
IC_{map}	0.04	0.02	0.01	0.01	0.01

Table I. Average error for (7) as estimate of the IC mean and mode, respectively, at different values for the joint count n_{11}.

We propose the the following m.a.p. estimate:

$$IC_{map} \approx \log_2 \frac{E[p_{11}]}{E[p_{1\cdot}]E[p_{\cdot 1}]} \tag{7}$$

The same expression has been used earlier as a crude estimate for the IC mean [7, 8]. To study the accuracy of this expression as an estimate for on one hand the mean and on the other hand the mode of the IC distribution, estimated values were compared to Monte Carlo simulated values based on 50,000 draws from each posterior IC distribution (the mode of the simulated IC distributions was estimated based on the empirical relationship $mode \approx 3 \cdot median - 2 \cdot mean$ for unimodal curves of moderate asymmetry [30]). A random subset of 1,000 drug-ADR combinations that occur in the WHO database were used for evaluation. Throughout, the moderating prior distribution described in Section 2.1 was used. The results are displayed in Figure 1. Clearly, (7) is a better estimate of the mode than of the mean.

2.3. Improved IC credibility interval estimates

Denote by IC_{025}, the 2.5 percentile of the posterior IC distribution. This is the lower limit of a two-sided 95% credibility interval for the IC, by which associations are typically ranked in IC analysis [7, 8]. The use of a lower credibility interval limit accounts for uncertainty in a conservative manner. The idea is to choose an estimate so that the true value is greater than the estimate with a given degree of certainty (here 97.5%). Together with the moderating prior distribution (see Section 2.1) this helps to reduce the number of false leads generated by IC analysis.

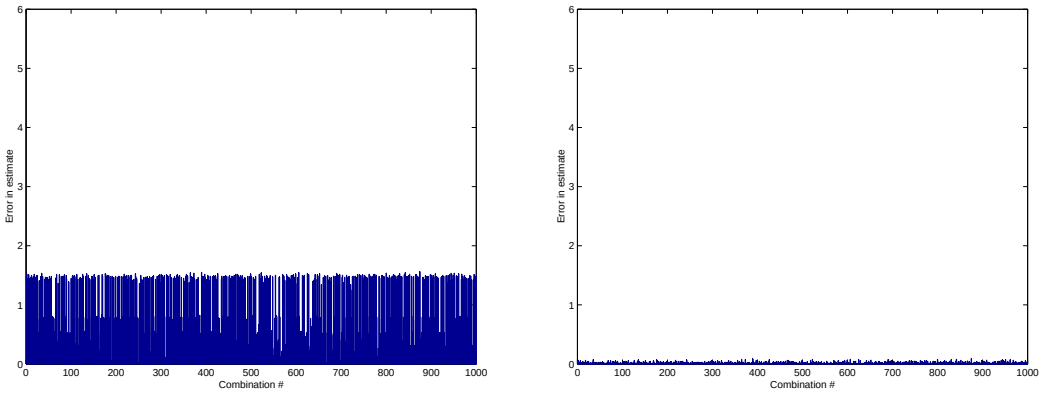
The IC credibility interval estimates were previously based on a normal approximation for the IC distribution [7, 8]. Monte Carlo experiments indicate that while the IC distribution tends to a normal distribution asymptotically (for large n_{11}), the assumption leads to a rather crude approximation for rare pairs of events ($n_{11} \leq 10$). Since more than 80% of the observed drug-ADR pairs in the WHO database fall into this critical category, the need for improvement is clear [27]. The use of brute force Monte Carlo simulation to estimate the posterior percentiles would give arbitrarily accurate estimates, but at too high a cost in computational complexity. Instead, we propose an approach based on an approximate formula for the difference between the mode and the lower credibility interval limit for the IC distribution.

Let Δ_{025} denote the true difference between IC_{map} and IC_{025}. Given estimates for IC_{map} and Δ_{025}, IC_{025} can be estimated as follows:

$$\hat{IC}_{025} = \hat{IC}_{map} - \hat{\Delta}_{025} \tag{8}$$

Empirical testing suggests that functions of the following general form model Δ_{025} well (A_r and B_r are fitted parameters):

$$\Delta_{025}(\gamma_{11}) = A_r \cdot \gamma_{11}^{-1/2} + B_r \cdot \gamma_{11}^{-3/2} \tag{9}$$



a. Normal approximation b. Proposed approximation

Figure 2. Errors in IC_{025} estimates. This figure shows deviations from Monte Carlo simulated values of the two IC_{025} estimates for 1,000 randomly selected drug-ADR pairs in the WHO database. 50,000 draws were used in each Monte Carlo simulation.

	$n_{11}=1$	$n_{11}=2$	$n_{11}=3$	$n_{11}=4$	$n_{11}=5$
Normal approximation	1.47	0.78	0.52	0.38	0.31
Proposed approximation	0.06	0.07	0.06	0.04	0.04

Table II. Average error in the IC_{025} estimates for different values of the joint count n_{11} among the 1,000 drug-ADR pairs from the WHO database.

In particular:

$$\lim_{\gamma_{11}\to\infty} \Delta_{025}(\gamma_{11}) = 0 \tag{10}$$

and because there are only 2 fitted parameters, there is little risk for over-fitting.

By using different parameters A_r and B_r for different ratios $r = \gamma_{11}/\min(\gamma_{1\cdot}, \gamma_{\cdot 1})$, the impact of the smaller of the two marginal parameters may be accounted for. We have estimated constants A_r and B_r for 11 different values of r $(0.0, 0.1, \ldots, 1.0)$, and use linear interpolation in between. Thus, for a given ratio r, the value of Δ_{025} is estimated by the weighted average of the Δ_{025} values for the two closest values of r for which there are fitted constants A_r and B_r available. For details about fitting the parameters in (9) and their values for different r, see Appendix I.1.

To evaluate the accuracy of the proposed approach to estimate IC_{025}, we compared Monte Carlo simulated values based on 50,000 draws to the estimated values, for the same data set as in Section 2.2. For comparison, the accuracy of the normal approximation [7, 8] was also evaluated. The results are displayed in Figure 2. Clearly, the proposed approximation is more accurate.

2.4. Stratified IC analysis

Although the purpose of dependency derivation is hypothesis generation, and a certain number of false leads is acceptable in this context, it is important to keep the proportion of false

leads at a minimum. One approach to improving the specificity is to detect and control for potential confounders. As for other epidemiological applications, adjusted overall estimates may be quoted when there is no suspicion of effect modification, otherwise stratum specific estimates should be used [31]. There are at least two different ways to adjust the IC for potential confounders.

The most obvious adjusted IC estimate is a weighted average of the stratum specific IC values as previously suggested [32]. However, this approach requires a careful selection of stratification variables, because it is particularly sensitive to data thinning. Strata with few or no observations of the event combination of interest will yield unreliable stratum specific IC estimates, and since the weights in calculating the pooled estimate are not necessarily correlated to the reliability of the estimates, this may lead to an unreliable adjusted IC estimate. Indeed, tentative experiments based on Monte Carlo simulation indicate that this approach to adjusting the IC typically leads to wider credibility intervals than for the unadjusted IC, which, in addition to the loss in precision, is a technical disadvantage since it makes more difficult the derivation of accurate credibility intervals.

An alternative approach to calculating pooled IC estimates is to use a Mantel-Haenszel type of adjustment where the denominator in the IC_{map} formula is equal to the weighted average of the expected joint probabilities in the different strata:

$$IC_{map} \approx \log_2 \frac{E[p_{11}]}{\sum_{k=1}^{n} E[p_{1\cdot|k}]E[p_{\cdot 1|k}] \cdot E[p_{\cdot\cdot k}]} \tag{11}$$

For this adjustment of the IC value, the moderating prior distribution may be used for both the numerator and for each term in the denominator. Since the numerator is not affected by the adjustment, it seems likely that the spread for the pooled IC should be the same as that for the unadjusted IC, and empirical testing supports this assumption. This indicates that the approximate credibility intervals proposed in Section 2.3 may be used for the adjusted IC_{025} as well.

As an illustration of the general usefulness of stratified IC analysis, we have investigated the association in the WHO database between the terms *sudden infant death syndrome* and *Polio virus vaccine live oral*. The unadjusted IC_{map} estimate for this association is 4.78 and the corresponding IC_{025} estimate is 4.63. However, since the Polio virus vaccine is typically given to small children and only small children suffer from SIDS (the sudden infant death syndrome), this is likely to be confounded by age [9].

There are 7 predefined age groups in the WHO database: *unspecified*, *0 - 1 month*, *2 months - 4 years*, *5 - 11 years*, *12 - 16 years*, *17 - 69 years* and *70+ years*. Table III displays stratum specific IC values for the association between SIDS and the Polio virus vaccine for these age groups. Based on this stratification, the adjusted IC_{map} estimate according to (11), is 1.19 and the corresponding IC_{025} estimate is 1.00. Clearly the stratification by age reduces the apparent strength of association. At the same time, the relatively strong association between SIDS and the Polio virus vaccine in the *age: unspecified* stratum renders dubious the listing of any overall IC estimate (adjusted or not). In this situation, a list of stratum specific IC values is probably a more appropriate output. Please note that a proper examination of this quantitative association would require the consideration of other potential confounders as well.

Some problems with routine stratification by a limited set of predefined variables have been pointed out previously [33]. For the WHO database, we use association specific stratification in the post-processing of clinically interesting drug-ADR pairs.

Stratum	$n_{polio,sids}$	$n_{polio\cdot}$	$n_{\cdot sids}$	$n_{\cdot\cdot}$	$IC_{polio,sids}$
unspecified	25	1126	87	572573	5.25 (4.64)
0 - 1 month	29	1408	79	9066	1.21 (0.73)
2 months - 4 years	203	30068	508	155209	1.04 (0.87)
5 - 11 years	0	5232	3	80140	-0.48 (-11.10)
12 - 16 years	0	299	0	63911	0.00 (-10.65)
17 - 69 years	0	461	13	1669422	-0.01 (-10.67)
70+ years	0	10	0	453481	0.00 (-10.66)

Table III. Stratum specific IC values for the association between SIDS and the Polio virus vaccine in different age groups. The numbers listed in the rightmost column are IC_{map} estimates with the moderating prior (IC_{025} estimates in brackets).

IC_{025}	Old +	Old -
New +	80363	616
New -	3532	522707

Table IV. A cross-classification of the observed drug-ADR pairs in the WHO database, with respect to the signs of the IC_{025} values for the two methods.

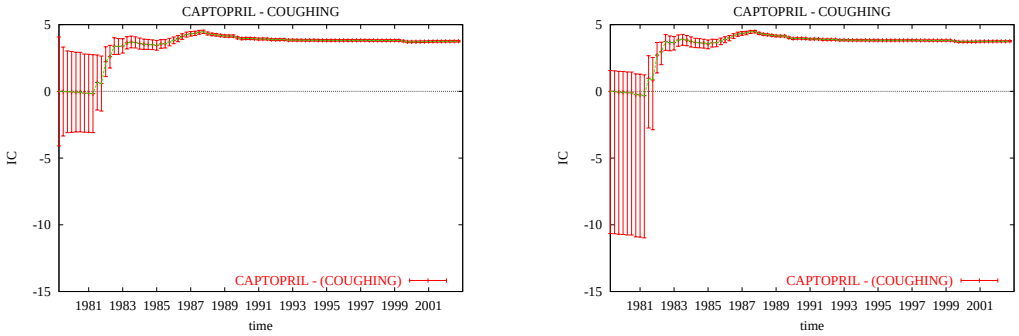
2.5. Example: a scan for drug-ADR associations in the WHO database

To study in practise, the impact of the proposed changes to IC analysis (new prior distribution and improved credibility interval estimates). We have carried out a complete scan of the WHO database (as of quarter 3, 2003) with both methods.

Table IV displays a cross-classification of all observed drug-ADR pairs in the WHO database with respect to whether the IC_{025} values are positive or negative (this is the threshold used in routine screening of the WHO ADR database) with the old and the new approach respectively. Clearly, the agreement between the two approaches is quite good: with respect to this threshold, the two methods differ for only around 4,000 out of the close to 600,000 observed drug-ADR pairs and Cohen's kappa measure is 0.97 (a Cohen's kappa of 1 would indicate perfect agreement). Where the two methods differ, the new approach seems to be somewhat more conservative, but there are event pairs for which the new but not the old IC_{025} estimate exceeds 0. These tend to have low joint counts n_{11} (ranging from 3 to 12) and low marginal counts $n_{1\cdot}$ for the drug (ranging from 3 to 68 in all but three cases, for which the values are significantly higher). In particular, the new approach alone highlights 84 event pairs where there is 3 reports in total for the drug – all on the same reaction. Because these event pairs may correspond to important problems for recently marketed drugs, it is a strength from a monitoring perspective that the new approach highlights them.

2.6. Example: a captopril-coughing time scan

To further examine the practical impact of the proposed changes to IC analysis, we studied the evolution in time of the IC between the drug substance captopril and the ADR term coughing with the two approaches. The association between captopril and coughing has been well known since 1986, but earlier work has shown that if IC analysis had been in use at the



a. Old approach b. New approach

Figure 3. The evolution over time of the IC between captopril and coughing, with the old and the new approach. Central estimates together with 95% credibility intervals are marked in the plot.

time a quantitative association between captopril and coughing would have been highlighted already in 1981 [7]. We were interested to see whether the proposed changes to IC analysis would delay or expedite the highlighting of this quantitative association in the database.

Figure 3 displays the change over time for the IC central estimates (together with 95% credibility interval estimates) for the captopril-coughing quantitative association, based on the old and new IC analysis approaches. The credibility interval estimates differ when the joint count is low, but the association would be highlighted in the same quarter regardless of which approach was used.

3. IC analysis for higher order dependency derivation

The IC as defined in Section 2 is a strength of association measure for pairs of events only, but there is often an interest in higher order associations. In the drug safety application, this may include drug-drug interactions or three way associations involving a drug substance, an ADR term and another risk factor (*e.g.* age or gender). Generally, a higher order strength of association measure should capture disproportionality in the occurrence of groups of events in the data set, which is not explicable by lower order associations. For three way associations, we would be interested in sets of three events that occur unexpectedly often even when pairwise associations between the events are accounted for.

The usefulness of extending the IC value to higher order associations is not necessarily limited to the dependency derivation application. Higher order IC values could be introduced in both the feedforward and the recurrent BCPNN in order to improve performance in classification and unsupervised pattern recognition, respectively.

An extension of the IC to third order associations was previously proposed in [8], but this did not compensate for pairwise associations. We, instead, propose the following definition of the third order IC:

$$IC_{xyz} = IC_{xy|z} - IC_{xy} \tag{12}$$

where:

$$IC_{xy|z} = \log_2 \frac{P(x, y \mid z)}{P(x \mid z)P(y \mid z)} \tag{13}$$

The logic behind this definition of the third order IC is that if there is a positive third order association, the presence of the third event should make the pairwise association between the other two events stronger (*i.e.* $IC_{xy|z}$ should exceed IC_{xy} etc). Conversely, if there is a negative third order associations, the presence of the third event should make the pairwise association between the other two events weaker.

It is easy to show that this definition is symmetric in x, y and z:

$$\begin{aligned} IC_{xyz} &= IC_{xy|z} - IC_{xy} = \\ &= IC_{xz|y} - IC_{xz} = \\ &= IC_{yz|x} - IC_{yz} \end{aligned} \tag{14}$$

since with simple algebraic operations, we can re-express IC_{xyz} as:

$$\begin{aligned} IC_{xyz} &= \log_2 \frac{P(y, z \mid x)}{P(y \mid x)P(z \mid x)} - \log_2 \frac{P(y)P(z)}{P(y, z)} = \\ &= \log_2 \frac{P(x, y, z)P(x)P(y)P(z)}{P(x, y)P(x, z)P(y, z)} \end{aligned} \tag{15}$$

The third order IC can be seen as an observed-to-expected ratio, where the expected value accounts for both main effects and pairwise interactions. To see this, let:

$$W_{x_1 \ldots x_n} = \frac{P(x_1, \ldots, x_n)}{P(x_1) \cdot \ldots \cdot P(x_n)}$$

Then the third order IC_{xyz} may be re-expressed as:

$$IC_{xyz} = \log_2 \frac{P(x, y, z)}{P(x)P(y)P(z)W_{xy}W_{xz}W_{yz}} \tag{16}$$

which is an approximate observed to expected ratio accounting for pairwise associations as well as marginal probabilities.

The generalisation of the IC to even higher orders is straightforward. For example, the fourth order IC can be defined as follows:

$$\begin{aligned} IC_{xyzv} &= IC_{xyz|v} - IC_{xyz} = IC_{xyv|z} - IC_{xyv} = \\ &= IC_{xvz|y} - IC_{xvz} = IC_{yzv|x} - IC_{yzv} \end{aligned} \tag{17}$$

which gives:

$$IC_{xyzv} = \ldots = \log_2 \frac{P(x, y, z, v)}{P(x)P(y)P(z)W_{xy}W_{xz}W_{xv}W_{yz}W_{yv}W_{zv}W_{xyz}W_{xyv}W_{xzv}W_{yzv}} \tag{18}$$

As desired, the approximate expected joint probability in the denominator accounts for both second and third order associations in addition to the marginal probabilities.

Most of the theory developed in Section 2 for pairwise IC values holds approximately for higher order IC values. A third order IC m.a.p. estimate similar to that for pairwise IC is:

$$IC_{map} \approx \log_2 \frac{E[p_{111}]E[p_{1\cdot\cdot}]E[p_{\cdot 1\cdot}]E[p_{\cdot\cdot 1}]}{E[p_{11\cdot}]E[p_{1\cdot 1}]E[p_{\cdot 11}]} \tag{19}$$

	n_{xyz}	$n_{xy\cdot}$	$n_{x\cdot z}$	$n_{\cdot yz}$	$n_{x\cdot\cdot}$	$n_{\cdot y\cdot}$	$n_{\cdot\cdot z}$	$n_{\cdots}$	IC_{xyz}
Drug: ketoconazole	5	63	27	11	6083	3695	5071	3176114	2.32 (1.08)
Age: 17-69 years	52	63	3764	2046	6083	3695	1756414	3176114	0.41 (0.20)
Gender: female	38	63	3427	1607	6083	3695	1753445	3176114	0.43 (0.09)
Country: USA	45	63	2718	2342	6083	3695	1478959	3176114	0.26 (-0.05)
Country: Germany	8	63	174	469	6083	3695	195102	3176114	0.94 (-0.18)

Table V. The top 5 third order IC values with terfenadine and ventricular fibrillation. The numbers listed are m.a.p. estimates with the moderating prior distribution (IC_{025} estimates in brackets).

Credibility intervals for third order IC values may be calculated with the formula proposed in Section 2.3 if, in the definition of r, $\min(\gamma_{1\cdot}, \gamma_{\cdot 1})$ is replaced by $\min(\gamma_{1\cdot\cdot}, \gamma_{\cdot 1\cdot}, \gamma_{\cdot\cdot 1})$. Adjustment of higher order IC values to control for confounders is also possible. For third order IC values, the Mantel-Haenszel adjusted m.a.p. estimate is:

$$IC_{map} \approx \log_2 \frac{E[p_{111}]}{\sum_{k=1}^{n} \frac{E[p_{11\cdot|k}]E[p_{1\cdot 1|k}]E[p_{\cdot 11|k}]}{E[p_{1\cdot\cdot|k}]E[p_{\cdot 1\cdot|k}]E[p_{\cdot\cdot 1|k}]} \cdot E[p_{\cdot\cdot k}]} \tag{20}$$

Furthermore, it is straightforward to generalise the moderating prior distribution described in Section 2.1 to higher order IC values (see Appendix I.2).

3.1. Example: A risk factors scan

Higher order IC analysis may be used to search for factors that influence the risk of a certain ADR given a particular drug. If, for example, the third order IC between a certain drug substance x, a certain ADR term y and a certain age group z were positive, this may indicate that patients of age group z are more prone to experiencing x-induced y than the population in general. Routine scans for third order IC values between a drug substance, an ADR term and some other factor (*e.g.* a certain gender or an age groups) may therefore be used to generate hypotheses with respect to potential high risk groups of patients. Positive higher order IC values may also be indicative of confounding, but for confounders, further investigation will show no significant variation in the IC values over the different strata.

Terfenadine was withdrawn due to concerns about its cardiotoxicity. Additionally, terfenadine and ketoconazole are known to interact so that the risk of heart problems is higher when the two are co-administered. Indeed, there are 5 reports on terfenadine, ketoconazole and ventricular fibrillation in the WHO database and the corresponding third order IC value is 2.32 with a lower credibility interval limit of 1.08. If we were to examine all three way associations between terfenadine and ventricular fibrillation and other events related to age, country, gender or other medication, there are 27 other events that occur at least once together with terfenadine and ventricular fibrillation on reports in the data set. Out of these, only 2 events other than the co-administration of ketoconazole have positive third order IC_{025} values with terfenadine and ventricular fibrillation (see Table V for the top 5 associations with respect to IC_{025} values). Based on this analysis, ketoconazole is clearly the most influential risk factor for this association.

4. Discussion

The analysis of spontaneous reporting data remains the cornerstone of post-marketing drug safety surveillance. Despite problems with data heterogeneity, it is the most important source of information for discovering previously unknown adverse effects from drugs after they are introduced on the market. *IC* analysis has proven to be an efficient method for exploratory quantitative analysis of post-marketing drug safety data [15] that while meeting the computational requirements also provides sophisticated protection against spurious associations. However, *IC* analysis as originally implemented [7, 8] is based on large sample approximations, and despite the large total number of reports in the WHO drug safety database the number of reports on a given drug-ADR pair is typically small (due to the large number of drug substances and ADR terms involved). Thus there is a clear need for the improved credibility intervals proposed in this article, and the results presented in Section 2.3 indicate that they do lead to improved accuracy and may allow for earlier discovery of problems related to recently marketed drug substances. These results are based on randomly selected drug-ADR pairs from the WHO database, but we expect the conclusions to hold generally for rare events in large and sparse data sets.

The Mantel-Haenszel adjustment for the *IC* proposed in Section 2.4 is important in that it will allow for robust exploratory data analysis in the presence of confounding. However, more research is needed to specify efficient strategies for how and when to carry out stratified analyses of spontaneous reporting data. It is, at present time, unclear whether routine adjustment by set of pre-defined variables for all event pairs in the database is to be preferred over unadjusted estimates in the initial screening of the database [33]. If higher order *IC* analysis or other sophisticated pattern recognition methods could be used for automated confounder detection, this may allow for data driven association specific adjustment by suspected confounders, and we aim to investigate this further in the future. The strong association between SIDS and the Polio vaccine in the *age unspecified* stratum of the WHO ADR database (see Section 2.4) is likely to be due to residual confounding and emphasises the problem of missing data for the stratification variables. This issue too needs to be resolved before optimal use of stratified dependency derivation is possible.

While the quantitative improvements for pairwise *IC* analysis proposed in Section 2 are refinements of the existing methodology, the generalisation to higher order associations in Section 3 allows for altogether new types of analysis related to complex quantitative associations. In combination with our methods for unsupervised pattern recognition [25], the methods presented in this article provide a comprehensive range of techniques for efficient knowledge discovery in spontaneous reporting data. An alternative approach to studying higher order associations would be to fit a generalised linear model with interaction terms, and in a similar spirit, other groups have proposed observed-to-expected ratios where the expected frequency is calculated based on a fitted log-linear model [10]. The advantage of higher order *IC* analysis in this context is that it is more direct (it does not require iterative methods for fitting) and allows for local analysis (in the sense that the higher order *IC* value for a certain set of events is only influenced by the joint and marginal counts for that specific set of events). Drug-drug interaction detection is a type of higher order association which is particularly important in the quantitative analysis of spontaneous reporting data and several approaches have been proposed [34, 35, 10]. In theory there is no obvious reason why higher order *IC* analysis could not be used to screen for drug interactions as well as any other risk factors,

but there has recently been a tendency to focus on more simple methods for the detection of drug-drug interactions [36], which indicates that more research into the basic characteristics of drug-drug interactions spontaneous reporting may be needed to resolve this issue successfully.

5. Conclusions

Earlier, IC analysis has proven useful in hypothesis generation with respect to quantitative associations in large drug safety data sets. In this article we have proposed improved methods for posterior inference in IC analysis, including an accurate estimate for the mode and significantly improved credibility interval estimates. In addition, we have extended the IC strength of association measure to higher order associations and illustrated the usefulness of this on real world data. An adjustment of the IC to control for potential confounders has also been described and applied to real world data.

REFERENCES

1. Evans SJ. Pharmacovigilance: a science or fielding emergencies? *Statistics in Medicine*, 2000. **19**(23):3199–3209.
2. Edwards IR. Spontaneous reporting–of what? Clinical concerns about drugs. *British Journal of Clinical Pharmacology* 1999; **48**(2):138–41.
3. Edwards IR. Spontaneous ADR reporting and drug safety signal induction in perspective. To honour Professor Jens Schou. *Pharmacology & Toxicology* 2000; **86**(s1):16–19.
4. Rawlins MD. Spontaneous reporting of adverse drug reactions. I: the data. *British Journal of Clinical Pharmacology* 1988; **26**(1):1–5.
5. Rawlins MD. Spontaneous reporting of adverse drug reactions. II: Uses. *British Journal of Clinical Pharmacology* 1988; **26**(1):7–11.
6. Edwards IR, Olsson S. WHO Programme - global monitoring. In *Pharmacovigilance*, Mann RD, Andrews EB (eds). Wiley:Chichester, 2002; 169–182.
7. Bate A, Lindquist M, Edwards IR, Olsson S, Orre R, Lansner A, De Freitas RM. A Bayesian neural network method for adverse drug reaction signal generation. *European Journal for Clinical Pharmacology* 1998; **54**:315–321.
8. Orre R, Lansner A, Bate A, Lindquist M. Bayesian neural networks with confidence estimations applied to data mining. *Computational Statistics & Data Analysis* 2000; **34**:473–493.
9. DuMouchel W. Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting systems. *American Statistician*, 1999; **54**:177–202.
10. DuMouchel W, Pregibon D. Empirical Bayes screening for multi-item associations. In *Knowledge Discovery and Data Mining*, 2001; 67–76.
11. Evans SJ, Waller PC, Davis S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiology and Drug Safety*, 2001. **10**(6):483–486.
12. van der Heijden PG, van Puijenbroek EP, van Buuren S, van der Hofstede JW. On the assessment of adverse drug reactions from spontaneous reporting systems: the influence of under-reporting on odds ratios. *Statistics in Medicine*, 2002; **21**(14):2027–2044.
13. Norwood PK, Sampson AR. A statistical methodology for postmarketing surveillance of adverse drug reaction reports. *Statistics in Medicine*, 1988. **7**(10):1023–1030.
14. Praus M, Schindel F, Fescharek R, Schwarz S. Alert systems for post-marketing surveillance of adverse drug reactions. *Statistics in Medicine*, 1993. **12**(24):2383–2393.
15. Lindquist M, Ståhl M, Bate A, Edwards IR and Meyboom RH. A retrospective evaluation of a data mining approach to aid finding new adverse drug reaction signals in the WHO international database. *Drug Safety* 2000; **23**(6):533–542.
16. Lindquist M, Edwards IR, Bate A, Fucik H, Nunes AM, Ståhl M. From association to alert - a revised approach to international signal analysis. *Pharmacoepidemiology and Drug Safety* 1999; **8**:15–25.
17. Bate A, Lindquist M, Orre R, Edwards IR, Meyboom RH. Data-mining analyses of pharmacovigilance signals in relation to relevant comparison drugs. *European Journal for Clinical Pharmacology*, 2002. **58**(7):483–490.

18. Ståhl M, Lindquist M, Edwards IR, Brown EG. Introducing triage logic as a new strategy for the detection of signals in the WHO drug monitoring database. *Drug Safety*, In Press.
19. Ståhl M, Edwards IR, Bowring G, Kiuru A, Lindquist M. The usefulness and use of signals from the WHO database by national pharmacovigilance centres - results from a questionnaire. *Drug Safety*, In Press.
20. Coulter DM, Bate A, Meyboom RH, Lindquist M, Edwards IR. Antipsychotic drugs and heart muscle disorder in international pharmacovigilance: data mining study. *British Medical Journal* 2001. **322**(7296): 1207–1209.
21. Sanz EJ, De-las-Cuevas C, Kiuru A, Bate A, Edwards IR. Paroxetine in pregnant women: SSRI and Neonatal withdrawal syndrome. *The Lancet* **In press**
22. Lansner A, Ekeberg Ö. A one-layer feedback artificial neural network with a Bayesian learning rule. *International Journal of Neural Systems* 1989; **1**:77–87.
23. Holst A, Lansner A. A higher order Bayesian neural network for classification and diagnosis. In *Applied Decision Technologies: Computational Learning and Probabilistic Reasoning*; Gammerman A (ed). Wiley: New York, 1996; 251–260.
24. Orre R, Lansner A. Pulp quality modelling using Bayesian mixture density neural networks. *Journal of Systems Engineering* 1996; **6**:128–136.
25. Orre R, Bate A, Norén GN, Swahn E, Arnborg S, Edwards IR. A Bayesian recurrent neural network for unsupervised pattern recognition in large incomplete data sets. In *On Data Mining and Classification Using a Bayesian Confidence Propagation Neural Network,* Orre R. Royal Institute of Technology, Sweden, 2003.
26. Norén GN, Orre R. Case based imprecision estimates for Bayes classifiers with the Bayesian bootstrap. *Machine Learning*; **In press**
27. Norén N. A Monte Carlo method for Bayesian dependency derivation. Master's thesis, Chalmers University of Technology, 2002.
28. Gelman A, Carlin JB, Stern HS, Rubin DB *Bayesian Data Analysis* (1st edn). Chapman & Hall: 1995.
29. Koski T, Orre R. Statistics of the Information Component in Bayesian neural networks. Technical report, Department of Numerical Analysis and Computing Science, Royal Institute of Technology, Stockholm, Sweden, 1998.
30. Kenney JF, Keeping ES *Mathematics of Statistics, Pt 1* (3rd edn). Van Nostrand: 1962; 50–54.
31. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 1959; **22**:719–748.
32. Gould L. Practical pharmacovigilance analysis strategies. *Pharmacoepidemiology and Drug Safety*, 2003; **12**:559–574.
33. Bate A, Edwards IR, Lindquist M, Orre R. Violation of Homogeneity: A Methodological Issue in the Use of Data Mining Tools. The authors' reply. *Drug Safety*, 2003; **26**(5):363–366.
34. Amery WK. Post-marketing drug safety management: a pharmaceutical industry perspective. *International Journal of Risk and Safety in Medicine*, 1994; **5**:67–270.
35. van Puijenbroek EP, Egberts ACG, Meyboom RHB, Leufkens HGM. Signalling possible drug-drug interactions in a spontaneous reporting system: delay of withdrawal bleeding during concomitant use of oral contraceptives and itraconazole. *British Journal of Clinical Pharmacology*, 1999; **47**:689–693.
36. Yang X, Fram DM. Using disproportional analysis as a tool to explore severe drug-drug interactions in AERS database. *Pharmacoepidemiology and Drug Safety*, 2004; **13**1:S247.

APPENDIX

I.1. Δ_{025} parameter fitting

Constants A_r and B_r for 11 different values of r $(0.0, 0.1, \ldots, 0.9, 1.0)$ were fitted to Equation 9 based on simulated Δ_{025} values where γ_{11} ranged from 1 to 100, $\gamma_{1\cdot} = \gamma_{11}/r$, $\gamma_{\cdot 1} = 100,000$ and $\gamma_{\cdot\cdot} = 10,000,000$. Each simulated Δ_{025} value was based on 100 000 Monte Carlo draws from the posterior IC distribution of interest. Table VI displays the fitted constants for different values of r (in the parameter fitting, $r = 0$ was approximated by $r = 0.001$ and $r = 1$ was approximated by $r = 0.999$ for computational stability).

r	A_r	B_r
0.0	3.09	2.22
0.1	2.93	2.27
0.2	2.78	2.26
0.3	2.62	2.25
0.4	2.45	2.15
0.5	2.25	2.12
0.6	2.03	2.05
0.7	1.79	1.93
0.8	1.61	1.89
0.9	1.13	1.15
1.0	0.073	-0.081

Table VI. Fitted parameters for the Δ_{025} function for different values of r

I.2. Moderating prior for the third order IC

The hyper parameters of the moderating prior for third order IC values is:

$$\begin{aligned}
\alpha_{111} &= \frac{q_{11\cdot}q_{1\cdot 1}q_{\cdot 11}}{q_{1\cdot\cdot}q_{\cdot 1\cdot}q_{\cdot\cdot 1}} \cdot \alpha_{\ldots} & \alpha_{011} &= \frac{q_{01\cdot}q_{0\cdot 1}q_{\cdot 11}}{q_{0\cdot\cdot}q_{\cdot 1\cdot}q_{\cdot\cdot 1}} \cdot \alpha_{\ldots} \\
\alpha_{110} &= \frac{q_{11\cdot}q_{1\cdot 0}q_{\cdot 10}}{q_{1\cdot\cdot}q_{\cdot 1\cdot}q_{\cdot\cdot 0}} \cdot \alpha_{\ldots} & \alpha_{010} &= \frac{q_{01\cdot}q_{0\cdot 0}q_{\cdot 10}}{q_{0\cdot\cdot}q_{\cdot 1\cdot}q_{\cdot\cdot 0}} \cdot \alpha_{\ldots} \\
\alpha_{101} &= \frac{q_{10\cdot}q_{1\cdot 1}q_{\cdot 01}}{q_{1\cdot\cdot}q_{\cdot 0\cdot}q_{\cdot\cdot 1}} \cdot \alpha_{\ldots} & \alpha_{001} &= \frac{q_{00\cdot}q_{0\cdot 1}q_{\cdot 01}}{q_{0\cdot\cdot}q_{\cdot 0\cdot}q_{\cdot\cdot 1}} \cdot \alpha_{\ldots} \\
\alpha_{100} &= \frac{q_{10\cdot}q_{1\cdot 0}q_{\cdot 00}}{q_{1\cdot\cdot}q_{\cdot 0\cdot}q_{\cdot\cdot 0}} \cdot \alpha_{\ldots} & \alpha_{000} &= \frac{q_{00\cdot}q_{0\cdot 0}q_{\cdot 00}}{q_{0\cdot\cdot}q_{\cdot 0\cdot}q_{\cdot\cdot 0}} \cdot \alpha_{\ldots}
\end{aligned} \tag{21}$$

where:

$$\alpha_{\ldots} = 0.5 \cdot \frac{q_{1\cdot\cdot}q_{\cdot 1\cdot}q_{\cdot\cdot 1}}{q_{11\cdot}q_{1\cdot 1}q_{\cdot 11}} \tag{22}$$

and:

$$
\begin{array}{lll}
q_{1\cdot\cdot} = \dfrac{n_{1\cdot\cdot} + 1/2}{n_{\cdot\cdot} + 1} & q_{11\cdot} = \dfrac{n_{11\cdot} + 1/4}{n_{\cdot\cdot} + 1} & q_{0\cdot 1} = \dfrac{n_{0\cdot 1} + 1/4}{n_{\cdot\cdot} + 1} \\
q_{0\cdot\cdot} = \dfrac{n_{0\cdot\cdot} + 1/2}{n_{\cdot\cdot} + 1} & q_{10\cdot} = \dfrac{n_{10\cdot} + 1/4}{n_{\cdot\cdot} + 1} & q_{0\cdot 0} = \dfrac{n_{0\cdot 0} + 1/4}{n_{\cdot\cdot} + 1} \\
q_{\cdot 1\cdot} = \dfrac{n_{\cdot 1\cdot} + 1/2}{n_{\cdot\cdot} + 1} & q_{01\cdot} = \dfrac{n_{01\cdot} + 1/4}{n_{\cdot\cdot} + 1} & q_{\cdot 11} = \dfrac{n_{\cdot 11} + 1/4}{n_{\cdot\cdot} + 1} \\
q_{\cdot 0\cdot} = \dfrac{n_{\cdot 0\cdot} + 1/2}{n_{\cdot\cdot} + 1} & q_{00\cdot} = \dfrac{n_{00\cdot} + 1/4}{n_{\cdot\cdot} + 1} & q_{\cdot 10} = \dfrac{n_{\cdot 10} + 1/4}{n_{\cdot\cdot} + 1} \\
q_{\cdot\cdot 1} = \dfrac{n_{\cdot\cdot 1} + 1/2}{n_{\cdot\cdot} + 1} & q_{1\cdot 1} = \dfrac{n_{1\cdot 1} + 1/4}{n_{\cdot\cdot} + 1} & q_{\cdot 01} = \dfrac{n_{\cdot 01} + 1/4}{n_{\cdot\cdot} + 1} \\
q_{\cdot\cdot 0} = \dfrac{n_{\cdot\cdot 0} + 1/2}{n_{\cdot\cdot} + 1} & q_{1\cdot 0} = \dfrac{n_{1\cdot 0} + 1/4}{n_{\cdot\cdot} + 1} & q_{\cdot 00} = \dfrac{n_{\cdot 00} + 1/4}{n_{\cdot\cdot} + 1}
\end{array}
\qquad (23)
$$