Mathematical Statistics
Stockholm University

A statistical methodology for drug–drug interaction surveillance

G. Niklas Norén, Rolf Sundberg, Andrew Bate,
I. Ralph Edwards

Research Report 2007:6

ISSN 1650-0377

Postal address:

Mathematical Statistics
Dept. of Mathematics
Stockholm University
SE-106 91 Stockholm
Sweden

Internet:

http://www.math.su.se/matstat

Mathematical Statistics
Stockholm University
Research Report **2007:6**,
http://www.math.su.se/matstat

A statistical methodology for drug–drug interaction surveillance

G. Niklas Norén, Rolf Sundberg*, Andrew Bate,
I. Ralph Edwards

February 2007

Abstract

Interaction between drug substances may yield excessive risk for adverse drug reactions (ADRs) when two drugs are taken in combination. Collections of individual case safety reports (ICSR) related to suspected ADR incidents in clinical practice have proven very useful in post-marketing surveillance for pairwise drug–ADR associations, but have yet to reach their full potential for drug–drug interaction surveillance. In this paper, we implement and evaluate a shrinkage observed-to-expected ratio for exploratory analysis of suspected drug–drug interaction in ICSR data, based on comparison with an additive risk model. We argue that the limited success of previously proposed methods for drug–drug interaction detection based on ICSR data may be due to an underlying assumption that absence of interaction is equivalent to having multiplicative risk factors. We provide empirical examples of established drug–drug interaction highlighted with our proposed approach, that go undetected with logistic regression. A database wide screen for suspected drug–drug interaction in the entire WHO database is carried out to demonstrate the feasibility of the proposed approach. As always in the analysis of ICSRs, the clinical validity of hypotheses raised with the proposed method must be further reviewed and evaluated by subject matter experts.

KEY WORDS: Adverse drug reaction, exploratory analysis, interaction, shrinkage, surveillance.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91 Stockholm, Sweden. E-mail: rolfs@math.su.se.

1 Background

Individual case safety reports (ICSR) on suspected adverse drug reaction (ADR) incidents in clinical practice, otherwise known as spontaneous reports, remain the main source of information to detect unknown adverse reactions to drug substances that are already on the market (Rawlins 1988). While randomized clinical trials (RCT) identify a safety profile of a medicinal product before it is brought to market, some ADRs will first be detected in the large numbers of patients exposed in real world clinical practice. This is particularly true of ADRs that are rare or that occur only after extended periods of use. The Uppsala Monitoring Centre in Sweden maintains and analyses the world's largest collection of ICSRs (3.8 million reports from 1968 to 2006) on behalf of the WHO Programme for International Drug Monitoring. The pooling of ICSRs from different countries in an international database allows public health and patient safety issues to be detected earlier after drug launch than if based on only the analysis of national data sets (Olsson 1998).

ICSR data contains information only on those prescriptions of drugs that are believed to have lead to ADRs. There is no information on the total number of patients prescribed a certain drug, so absolute incidence or reporting rates are impossible to estimate. On account of their reliance on voluntary reporting, collections of ICSRs are open to potential reporting biases, such as the general under-reporting of known and less serious events and the relative over-reporting on drug–ADR combinations following attention in the scientific or public media. In addition, ICSRs entail problems with varying data quality (Edwards et al. 1990), the possible presence of duplicate case reports (Norén et al. 2005) and the vulnerability to intentional manipulation through fraudulent reporting (Stephens 2004). Still, ICSRs remain well accepted as the best data source currently available for the early detection of previously unsuspected ADRs.

The nature of ICSRs limits the strength of conclusions that can be drawn. Collections of ICSRs are unsuitable for hypothesis testing, but provide an important basis for hypothesis generation with the primary aim of highlighting potential public health or patient safety issues for further investigation (Bate et al. 1998). For large collections of ICSRs, quantitative methods are indispensable in screening the massive inflow of new reports (the WHO database currently receives over 200,000 new ICSRs each year). Automated knowledge discovery methods may also highlight interesting aspects of groups of ICSRs that are not immediately apparent in manual review.

1.1 Drug–drug interaction surveillance

The proportion of ADRs that are due to drug–drug interaction is thought to be between six and thirty per cent (Pirmohamed and Orme 1998). For

example, two drugs may compete for the same biologic receptor with a resulting antagonistic effect. Alternatively, one drug may inhibit an enzyme that metabolizes the other and thus cause ADRs due to an accidental overdose. Similarly, enzyme induction may lead to lack of effect of a co-medication, and this may also be considered as an ADR. A true drug–drug interaction is one where the pharmacological outcome is not just the direct result of the two drugs' individual effects (Pirmohamed and Orme 1998), and our interest is in effects that exceed that expected under simple independent action of each drug.

The early detection of ADRs due to suspected drug–drug interaction is important both from an overall public health perspective and the individual patient safety point of view. While many drug–drug interactions can be predicted based on pharmacological knowledge, ICSRs and other real world observational data provide an important complement, in particular for the detection of unpredictable drug–drug interaction. If previously unknown high risk drug combinations can be identified, they can potentially be avoided in the future, and if ADRs can be attributed to drug–drug interaction rather than to individual drugs, drugs that would have otherwise been withdrawn can remain on the market with warnings concerning co-medication.

ICSRs have a primarily structured format agreed internationally where the information related to the observed ADR incident can be entered. One or more drugs can be listed, at least one of which must be labelled as suspected of having caused the observed ADR. Co-administered drugs that the reporter considers to be unrelated to the observed ADR can be listed as such. The reporter can also list sets of drugs as specifically suspected of having interacted to cause the ADR. Other possibly useful information on reports includes dosage, therapy start and end dates and their relation to the onset date of the suspected ADR. There are also free text fields that may contain relevant pieces of information.

Even though reporters can explicitly list sets of drugs as suspected of having interacted, in many cases the drugs will be listed as co-suspected instead, or suspicion will even be aportioned to just one of the drugs. In order not to delay the early discovery of drug–drug interaction, surveillance should not focus solely on those ICSRs where the drugs are explicitly listed as suspected to interact. Similarly, there are free text fields that may in some instances allow clinical experts to draw conclusions about potential drug–drug interaction incidents based on single ICSRs, but as such information cannot be expected to be available generally, it is likely to be more useful in clinical review than for first pass screening purposes. In order to detect suspected drug–drug interaction as early as possible, we focus on the total number of reports on two drugs with a particular ADR, regardless of whether the two drugs are listed as suspected or interactive.

1.2 Statistical interaction

For the purpose of determining whether a high absolute reporting rate is indicative of interaction, statistical methodology is required. In statistical inference, interaction is defined in terms of departure from an additive model that accounts only for main effects. For example, an observed relative frequency p of a certain outcome under simultaneous exposure to X_1 and X_2 (indicator variables with observed values x_1 and x_2 equal to either 0 or 1), may be compared to the expected relative frequency under a no-interaction logistic regression model:

$$\log \frac{p}{1-p} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \tag{1}$$

Alternatively, the observed incidence rate λ may be compared to that expected under a no-interaction linear model:

$$\lambda = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 \tag{2}$$

Models (1) and (2) are clearly not equivalent, and since (1) is additive on the logit scale, its risk factors approximately multiply (provided p is not large), whereas (2) assumes additive incidence rates. The choice of baseline model will determine the nature and interpretation of an estimated interaction term. The two models define interaction as departure from their respective baseline assumptions, and sometimes, even the direction of estimated interaction may vary between models. Consider the potential interaction between two drugs with respect to a particular ADR. If, in the absence of the second drug, the risk for the ADR is 0.03 among patients exposed to the first drug vs 0.01 among patients not exposed to the first drug and that with the second drug co-prescribed, the corresponding risks are 0.10 and 0.05. Then the observed risk ratio for the first drug is higher in presence of the second drug (3.0 vs 2.0) whereas the risk difference is lower (+0.02 vs +0.05). The choice between a baseline model with additive or multiplicative risk factors thus determines the direction of the estimated interaction term.

Clearly, departure from a given statistical model does not automatically correspond to interesting interaction: the appropriate statistical baseline model depends on the subject matter question of interest. However, Rothman et al. (1980) argue that from both public health and individual patient safety perspectives, absolute differences in risk are more important than relative ones, and advise that interaction should ordinarily be defined in terms of departure from a model with additive risk factors. From a public health perspective, interaction relative to an additive risk model indicates whether the absolute number of cases in a population depends on to what extent two different risk factors co-occur. From an individual patient safety point of view, it indicates whether, for a given patient, the increase in absolute risk due to one risk factor is modified by the presence of the other. Interaction defined as departure

from a baseline model with additive risk factors thus provides a solid basis both for public health policy making and individual decisions (Rothman et al. 1980).

For the purpose of ADR surveillance, the additive risk baseline model has the advantage of estimating each drug's separate effect based on an absolute rather than a relative difference in relative reporting rates. In contrast, for a baseline model with multiplicative risk factors, if the relative reporting rate of the ADR is near 0 on reports that list neither of the two drugs of interest, even very modest relative reporting rates of the ADR for either drug on its own may yield considerable expected relative reporting rates of the ADR given co-prescription of the two drugs (because the ADR might still be many times more often reported given either drug than in the absence of both drugs). Moreover, when the background relative reporting rate of the ADR is very low, missing information on one of two interacting drugs will yield overestimated relative reporting rates for sole use of either drug. This will distort interaction estimates regardless of whether they are based on departure from baseline models with additive or multiplicative risk. However, in combination with very low background relative reporting rates, a more severe impact can be expected for baseline models with multiplicative risk.

1.3 Earlier work

Several methods have been proposed for quantitative drug–drug interaction detection in ICSR data sets. Most of the previously proposed methods have been based on departure from baseline models where risk factors essentially multiply. This is not surprising, given the general availability of such methods in standard software. Both van Puijenbroek et al. (1999) and van Puijenbroek et al. (2000) present interaction analyses based on logistic regression. DuMouchel and Pregibon (2001) propose an approach to interaction detection based on departure from a log-linear model. The higher order measure of disproportionality proposed in Norén et al. (2006) is also based on a no interaction model where risk factors multiply. In contrast, the methods for interaction detection proposed in Almenoff et al. (2003) and Yang and Fram (2004) compare the relative reporting rate of the ADR given co-prescription of two drugs, to the highest relative reporting rate of the ADR given sole prescription of either drug. Thus, they make no distinction between interaction and simple independent action, and are not appropriate for detecting drug–drug interaction as defined in the context of this paper.

No database wide screens for drug–drug interaction in ICSR data sets have been published and there are no reports in the literature suggesting that any of the proposed interaction detection methods have been implemented for routine ADR surveillance. Nor are we aware of any examples of early warnings on drug–drug interaction produced by any of these methods. DuMouchel and Pregibon (2001) present no empirical results for ADR data. and the empirical examples presented in the other papers tend to be on iso-

lated examples where the relative reporting rates for the ADR given sole prescription of either drug do not deviate considerably from the baseline relative reporting rate for the ADR in the absence of both drugs: 0 and 0.006 versus a background relative reporting rate of 0.005 in van Puijenbroek et al. (1999), 0.04 and 0.03 versus a background relative reporting rate of 0.03 in van Puijenbroek et al. (2000), and finally 0.004 and 0.002 versus a background relative reporting rate of 0.002 in Norén et al. (2006) (for further details, see Table 1 in Section 3). For these examples where the relative reporting rates given sole prescription of either drug are so close in magnitude to the background relative reporting rate in the absence of both drugs, the estimated separate effect of each drug will be very small and the choice of baseline model less critical.

1.4 Aim of this paper

The aim of this paper is to propose a disproportionality measure for exploratory analysis of suspected drug–drug interaction in ICSR data, starting from a baseline model with additive risk.

2 Method

In order to screen for disproportional reporting indicative of suspected drug–drug interaction in ICSR data, we formulate a model for the expected incidence of suspected ADRs in a population of interest and translate this to the context of the database. We compare the observed relative reporting rate f_{11} of an ADR given the co-prescription of two drugs in the database to its expected value $E[f_{11}]$ estimated from the relative reporting rates of the ADR given sole reporting of each drug, under the baseline assumption that the two drugs do not interact.

In the choice of absolute or relative difference between f_{11} and $E[f_{11}]$ as the basis for our measure of disproportionality, we consider the relative difference to be the more relevant measure, based on the view that for an interaction effect to be of interest it should represent a substantial proportion of the ADR incidents under consideration. As equivalent with the relative difference, we take as measure an observed-to-expected ratio analogous to that used in pairwise disproportionality analysis of ICSR data (Norén et al. 2006, DuMouchel and Pregibon 2001):

$$\frac{f_{11}}{E[f_{11}]} \tag{3}$$

While $E[f_{11}]$ is not known, it can be estimated, and f_{11} can be compared to this estimate.

2.1 Population model

We first model the occurrence in the population of the adverse event A of interest. New prescriptions occur under a certain average intensity that varies depending on the set of prescribed drugs. In connection with a given prescription, there is a certain risk (probability), dependent on the set of prescribed drugs, that the adverse event of interest (A) occurs and is reported as a suspected ADR. First, denote by α_0 the background risk for A due to for example progression of the underlying disease or a coincidental adverse event only temporally associated with the medical treatment. Next, consider two drugs D_1 and D_2, prescribed alone or in conjunction, or not at all. The total risk p_{00} for A in individuals who are prescribed neither D_1 or D_2 is just the background risk:

$$p_{00} = \alpha_0 \tag{4}$$

Let α_1 denote the risk for A attributable to D_1, and let α_2 denote the risk for A attributable to D_2. Under the assumption that the background risk of A, and the risks due to D_1 and D_2 are all mutually independent, the total risk p_{10} for A in individuals treated with D_1 in the absence of D_2 is:

$$\begin{aligned} p_{10} &= 1 - (1 - \alpha_0)(1 - \alpha_1) \\ &= \alpha_0 + \alpha_1 - \alpha_0 \cdot \alpha_1 \end{aligned} \tag{5}$$

Similarly, the total risk p_{01} for A in individuals treated with D_2 in the absence of D_1 is:

$$p_{01} = 1 - (1 - \alpha_0)(1 - \alpha_2) \tag{6}$$

The total risk p_{11} for A in individuals under combined treatment of D_1 and D_2 is:

$$p_{11} = 1 - (1 - \alpha_0)(1 - \alpha_1)(1 - \alpha_2) \tag{7}$$

Given that both the background risk, α_0, and the attributable risk from D_1, α_1, can be assumed to be small for any ADR A, their product, $\alpha_0 \cdot \alpha_1 << \alpha_0 , \alpha_1$. Thus, the following approximation of (5)is valid:

$$p_{10} \approx \alpha_0 + \alpha_1 \tag{8}$$

Similarly:

$$p_{01} \approx \alpha_0 + \alpha_2 \tag{9}$$

$$p_{11} \approx \alpha_0 + \alpha_1 + \alpha_2 \tag{10}$$

The absence of reliable information on the total number of different types of prescriptions as well as the degree of under-reporting, makes it difficult to

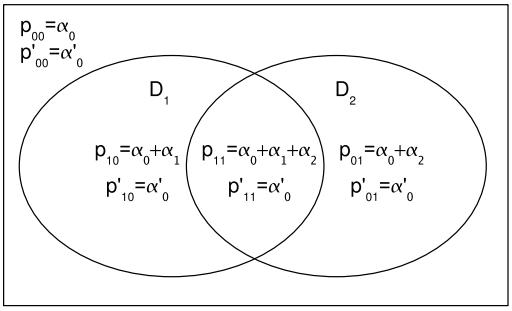

Figure 1: Venn diagram for the risks of A and A', in different subsets of the drug taking population.

link (4), (8), (9) and (10) directly to observed relative reporting rates in the database. In order to obtain a database reference related to the total number of prescriptions for different sets of drugs, let A' denote the occurrence of at least one of a (potentially large) group of ADRs excluding A (and in its absence so that A and A' are mutually exclusive events). Let α'_0 denote the background risk for A'. If ADRs with an attributable risk from either D_1 or D_2 can be excluded from A', the total risk for A' will be α'_0 for all possible combinations of D_1 and D_2:

$$\begin{aligned} p'_{00} &= \alpha'_0 \\ p'_{10} &= \alpha'_0 \\ p'_{01} &= \alpha'_0 \\ p'_{11} &= \alpha'_0 \end{aligned} \tag{11}$$

However, the identification of an appropriate set of unrelated ADR terms for a given pair of drugs requires expert clinical judgment, which cannot easily be automated for routine screening purposes. Common practice in pairwise disproportionality analysis of ADR surveillance data is therefore to include all ADRs other than A in A' for first pass screening purposes. We propose the same approach be used for interaction screening, since (11) will hold approximately unless D_1 or D_2 considerably alters the overall risk for any suspected ADR in association with the prescription. Should this be the case, restriction of A' to a more narrow set of ADRs will resolve the problem.

2.2 Database relative reporting rates

In order to obtain an estimator for the observed-to-expected ratio of the relative reporting rate in the database of A given D_1 and D_2 co-prescribed,

based on the population model in Section 2.1, let n_{111} denote the number of reports on A listing both D_1 and D_2, let n_{101} denote the number of reports on A listing D_1 but not D_2, let n_{011} denote the number of reports on A listing D_2 but not D_1 etc. Similarly, let $n_{1..}$ denote the total number of reports on D_1, $n_{.1.}$ the total number of reports on D_2 and $n_{..1}$ the total number of reports on A etc. Let:

$$\begin{aligned} f_{00} &= \frac{n_{001}}{n_{00\cdot}} \\ f_{10} &= \frac{n_{101}}{n_{10\cdot}} \\ f_{01} &= \frac{n_{011}}{n_{01\cdot}} \\ f_{11} &= \frac{n_{111}}{n_{11\cdot}} \end{aligned} \tag{12}$$

denote the corresponding observed relative reporting rates for A.

We will now construct an estimator for the expected relative reporting rate of A under combined use of D_1 and D_2 (f_{11}) based on the relative reporting rates of A given prescription of at most one of D_1 and D_2 (f_{00}, f_{10} and f_{01}). This will be the denominator of our observed-to-expected ratio in (3). In order not to let potential interaction contaminate the estimation of the expected relative reporting rate, we base it exclusively on f_{00}, f_{10} and f_{01}. Ignoring potential reporting biases, denote by r the probability that a suspected ADR incident is characterized as such by a health professional, reported to a pharmacovigilance center and eventually forwarded to the WHO programme (the impact of violations of this assumption of equal reporting rates is further discussed in Section 4). The expected value for the background relative reporting rate of A in the absence of both D_1 and D_2 is:

$$\begin{aligned} E[f_{00}] &= E[E[f_{00} \mid n_{00\cdot}]] \\ &= E\left[\frac{\alpha_0 \cdot r}{\alpha_0 \cdot r + \alpha_0' \cdot r}\right] \\ &= \frac{\alpha_0}{\alpha_0 + \alpha_0'} \end{aligned} \tag{13}$$

Similarly:

$$E[f_{10}] = \frac{\alpha_0 + \alpha_1}{\alpha_0 + \alpha_1 + \alpha_0'} \tag{14}$$

$$E[f_{01}] = \frac{\alpha_0 + \alpha_2}{\alpha_0 + \alpha_2 + \alpha_0'} \tag{15}$$

$$E[f_{11}] = \frac{\alpha_0 + \alpha_1 + \alpha_2}{\alpha_0 + \alpha_1 + \alpha_2 + \alpha_0'} \tag{16}$$

After re-expression of (16) in terms of (13–15):

$$\begin{aligned} E[f_{11}] &= \frac{\alpha_0 + \alpha_1 + \alpha_2}{\alpha_0 + \alpha_1 + \alpha_2 + \alpha_0'} \\ &= 1 - \frac{\alpha_0'}{\alpha_0 + \alpha_1 + \alpha_2 + \alpha_0'} \\ &= 1 - \frac{1}{\frac{\alpha_0+\alpha_1}{\alpha_0'} + \frac{\alpha_0+\alpha_2}{\alpha_0'} - \frac{\alpha_0}{\alpha_0'} + 1} \\ &= 1 - \frac{1}{\frac{E[f_{10}]}{1-E[f_{10}]} + \frac{E[f_{01}]}{1-E[f_{01}]} - \frac{E[f_{00}]}{1-E[f_{00}]} + 1} \end{aligned} \tag{17}$$

Thus, as estimator of $E[f_{11}]$, we may use:

$$g_{11} = 1 - \frac{1}{\frac{f_{10}}{1-f_{10}} + \frac{f_{01}}{1-f_{01}} - \frac{f_{00}}{1-f_{00}} + 1}$$

However, in order to avoid possible misleading influence of negative α_1 or α_2 estimates, we modify g_{11} as follows:

$$g_{11} = 1 - \frac{1}{\max\left(\frac{f_{00}}{1-f_{00}}, \frac{f_{10}}{1-f_{10}}\right) + \max\left(\frac{f_{00}}{1-f_{00}}, \frac{f_{01}}{1-f_{01}}\right) - \frac{f_{00}}{1-f_{00}} + 1} \tag{18}$$

When $f_{10} < f_{00}$ (indicating no risk for A attributable to D_1), this yields the most sensible estimator $g_{11} = \max(f_{00}, f_{01})$, and vice versa when $f_{01} < f_{00}$.

2.3 A shrunk interaction measure

To form a measure for the interaction seen in a data set we first consider:

$$\Omega_0 = \log_2 \frac{f_{11}}{g_{11}} \tag{19}$$

In spite of the very large data sets, the events involved in ADR surveillance should be rare, so g_{11} tends to be very small, and as a consequence Ω_0 is sensitive to spurious associations. This is a well known phenomenon in screening ICSR data sets for single drug–ADR excessive reporting rates, where the contingency tables are often extremely unbalanced. In that context, shrinkage has proven an effective approach to reduce the sensitivity to random fluctuations in measures of disproportionality, based on small amounts of data. Two of the most extensively used pairwise measures of disproportionality for ICSR data are indeed shrinkage measures: the Information Component (IC) (Bate et al. 1998) and the Empirical Bayes Geometric Mean (EBGM) (DuMouchel

and Pregibon 2001). Both of them are based on the pairwise observed-to-expected ratio of the relative reporting rate for an ADR together with a certain drug.

In order to construct a similar shrinkage measure from (19), we re-express the observed and expected relative reporting rates f_{11} and g_{11} in terms of the observed and expected counts n_{111} and $E_{111} = g_{11} \cdot n_{11\cdot}$:

$$\frac{f_{11}}{g_{11}} = \frac{n_{111}/n_{11\cdot}}{E_{111}/n_{11\cdot}} = \frac{n_{111}}{E_{111}} \tag{20}$$

and propose the Ω shrinkage measure:

$$\Omega = \log_2 \frac{n_{111} + \alpha}{E_{111} + \alpha} \tag{21}$$

Here, α is a tuning parameter determining shrinkage strength (higher α gives stronger shrinkage and vice versa). For $\alpha = 0$, we obtain Ω_0. The impact of α is equivalent to that of α additional expected reports on the ADR under joint prescription of the two drugs and lack of interaction, and an exactly matching increase in the observed count. Unlike in shrinkage regression, where tuning parameters can be selected on the basis of cross-validation estimates for classifier performance, there is no objective basis for choosing a particular value for α in disproportionality analysis. Empirical studies of the WHO database have indicated that $\alpha = 0.5$ provides just enough shrinkage to avoid the highlighting of case series consisting of less than 3 reports, and all subsequent Ω estimates presented in this paper are based on this value for the tuning parameter. However, other α-values may be more appropriate for ICSR data sets very different from the WHO database.

The Ω shrinkage measure can be motivated both from frequentist and Bayesian perspectives. In the frequentist perspective, Ω is biased towards 0 relative to Ω_0, but with better variance properties. As n_{111} and E_{111} increase, the difference between Ω and Ω_0 approaches 0. From the Bayesian perspective, Ω can be viewed as the logarithm of the posterior mean of an unknown rate of incidence μ under the natural assumption that n_{111} is $Po(\mu \cdot E_{111})$-distributed with $\log_2 \mu = \Omega$ and a gamma prior distribution (or random effects model in a likelihood-based analysis) for μ: $G(\alpha, \alpha)$, with expected value 1. The choice of prior is made mainly for mathematical convenience, since due to conjugacy the posterior distribution for μ will also be gamma (but with parameters $n_{111} + \alpha$ and $E_{111} + \alpha$, expected value $\frac{n_{111}+\alpha}{E_{111}+\alpha}$ and variance $\frac{n_{111}+\alpha}{(E_{111}+\alpha)^2}$).

With the Bayesian approach, exact credibility interval limits for μ can be found numerically as solutions to the following equation, for appropriate posterior quantiles μ_q:

$$\int_0^{\mu_q} \frac{(E_{111} + \alpha)^{n_{111}+\alpha}}{\Gamma(n_{111} + \alpha)} u^{n_{111}+\alpha-1} e^{-(n_{111}+\alpha)u}\, du = q \tag{22}$$

Specifically, the logarithm of the solutions to (22) for $q = 0.025$ and $q = 0.975$, respectively, provide the upper and lower limits of a two-sided 95% credibility interval for Ω: Ω_{025} and Ω_{975}.

In the frequentist approach, for large n_{111} and E_{111}, Ω differs little from Ω_0 and a Poisson (or binomial) confidence interval can be used. A crude estimator of the precision of Ω_0 based on the Poisson model is:

$$\begin{aligned} Var(\Omega_0) &= Var\left(\log_2 \frac{n_{111}}{E_{111}}\right) \\ &\approx Var(\log_2 n_{111}) \\ &= \frac{Var(\log n_{111})}{\log(2)^2} \\ &\approx \frac{Var(n_{111})}{n_{111}^2 \log(2)^2} \\ &\approx \frac{1}{n_{111} \log(2)^2} \end{aligned} \tag{23}$$

where, in the first approximation, any randomness in E_{111} has been assumed negligible.

3 Results

We carried out two investigations to study the usefulness of the proposed disproportionality measure Ω for drug–drug interaction detection. First, we compared Ω to a third order log odds ratio with respect to the ability to detect five examples of drug–drug interaction in ICSR data. Three of these were based on previously published studies of drug–drug interaction in ICSR data and two were examples of established drug–drug interaction based on WHO data. Second, we screened the entire WHO database for three-way disproportional reporting rates, to see whether the combinations of two drugs and one ADR with $\Omega_{025} > 0$ tend to be of clinical interest. This study also gave an indication as to the feasibility of using $\Omega_{025} > 0$ as a threshold for clinical review in screening ICSR data sets for suspected drug–drug interaction.

3.1 Case studies

We used data from three previously published studies of drug–drug interaction in ICSR data: delayed bleeding from concomitant use of itraconazole and oral contraceptives in van Puijenbroek et al. (1999), cardiac events from concomitant use of diuretics and NSAIDs in van Puijenbroek et al. (2000) and

ventricular fibrillation from concomitant use of terfenadine and ketoconazole in Norén et al. (2006). In addition, we considered two new examples of established drug–drug interaction with excessive relative reporting rates in the WHO database: "drug level increased" from concomitant use of digoxin and clarithromycin and rhabdomyolysis from concomitant use of cerivastatin and gemfibrozil. These examples were selected because an unpublished investigation based on the higher order IC for three-way disproportional reporting rates proposed in Norén et al. (2006), surprisingly indicated negative interaction for these two examples, despite the fact that they are well established examples of drug–drug interaction. In fact, co-prescription together with gemfibrozil was contraindicated for cerivastatin even as it was introduced on the market, and there are over a thousand case reports in the WHO database on rhabdomyolysis for concomitant use of cerivastatin and gemfibrozil. Moreover, as large a proportion as 75% of all case reports on cerivastatin together with gemfibrozil list rhabdomyolysis as (one of) the suspected ADR. This is to be compared with relative reporting rates of 0.1% in the absence of both cerivastatin and gemfibrozil, 4% for sole gemfibrozil use and 27% for sole cerivastatin use. Clearly, an interaction detection method which fails to highlight such reporting patterns as indicative of suspected drug–drug interaction will be of limited use in ADR surveillance.

Table 1 lists relevant data for all five examples considered in this investigation. Database counts (n_{111}, $n_{11\cdot}$, etc) for the first three case studies were taken directly from the corresponding publications. Data for the two new case studies was extracted from the WHO database as of 2004-12-31. For comparison, interaction terms (third order log-odds ratios) from a logistic regression model fitted directly to the database are provided for all five case studies (the estimates for the first two case studies correspond to those quoted in van Puijenbroek et al. (1999) and van Puijenbroek et al. (2000), respectively). Additionally, Ω and Ω_{025} values calculated according to the approach presented in Section 2.3 are provided for all five case studies.

The Ω measure of disproportionality indicates positive interaction for all five examples in Table 1. Disregarding shrinkage, each unit increase in Ω corresponds to a doubling of the observed-to-expected ratio. An Ω of 1 thus indicates that there are (at least – because of the shrinkage) twice as many reports on the ADR given the two drugs co-prescribed as we would expect, based on each drug's separate risk profile. However, as for any shrinkage measure, it is important not to over-interpret the specific value of Ω as it may depend strongly on the choice of tuning parameter α. Ω never exceeds the log observed-to-expected ratio, but if either n_{111} or E_{111} are small, the choice of α will determine to what extent Ω is shrunk towards 0. As is clear from a comparison with Ω_0 in Table 1, shrinkage has little impact on Ω in the second, fourth and fifth examples. As for the first and third examples where there are just 5 or 10 reports on the ADR given the two drugs of interest, the difference between Ω and Ω_0 is substantial.

An analysis based on logistic regression (odds ratios) indicates positive interaction for the examples in van Puijenbroek et al. (1999) and van Puijenbroek

Drug 1	Itraconazole	Diuretics	Terfenadine	Digoxin	Gemfibrozil
Drug 2	Oral Contraceptives	NSAIDs	Ketoconazole	Clarithromycin	Cerivastatin
ADR	Delayed bleeding	Cardiac events	Ventricular fibrillation	Drug level increased	Rhabdomyolysis
n_{111}	10	25	5	35	1084
$n_{11\cdot}$	23	278	27	85	1431
$n_{1\cdot 1}$	10	78	63	1193	1304
$n_{\cdot 11}$	19	67	11	245	3022
$n_{1\cdot\cdot}$	39	1775	6083	10650	6756
$n_{\cdot 1\cdot}$	1489	1613	5071	12390	9181
$n_{\cdot\cdot 1}$	39	305	3695	10781	6321
$n_{\cdots}$	5503	9822	$3.2 \cdot 10^6$	$3.2 \cdot 10^6$	$3.2 \cdot 10^6$
f_{00}	0.0050	0.028	0.0011	0.0030	0.001
f_{10}	0	0.035	0.0096	0.11	0.04
f_{01}	0.0061	0.031	0.0012	0.017	0.25
f_{11}	0.43	0.090	0.19	0.41	0.76
g_{11}	0.0061	0.039	0.0096	0.12	0.27
$\log_2(OR)$	$+\infty$	+1.23	+4.50	-0.03	-2.24
Ω_0	+6.15	+1.20	+4.27	+1.77	+1.47
Ω	+4.03	+1.16	+2.86	+1.72	+1.47
Ω_{025}	+3.00	+0.54	+1.33	+1.20	+1.38

Table 1: Empirical data for three case studies of suspected drug–drug interaction in ICSR data from the literature (van Puijenbroek et al. 1999, van Puijenbroek et al. 2000, Norén et al. 2006) together with data from the WHO database for two examples of established drug–drug interaction.

et al. (2000) (as already known from the original publications), as well as for that in Norén et al. (2006). In contrast, it fails to highlight examples 4 and 5 as indicative of suspected drug–drug interaction.

3.2 A database screen

As a complement to the investigation in Section 3.1 of whether disproportionality analysis based on Ω will highlight established examples of drug–drug interaction, we carried out a database wide screen for disproportional reporting in the entire WHO database. The aim of this investigation was to study to what extent the drug–drug–ADR combinations with $\Omega_{025} > 0$ in the WHO database correspond to clinically interesting suspected drug–drug interaction.

The presence of duplicate case reports is an important data quality problem that complicates knowledge discovery in ADR surveillance. In order to avoid problems with case report duplication in the analysis presented below, we pre-processed our extract from the WHO database (as of 2004-12-31) by completely removing any suspected duplicates highlighted by the duplicate detection algorithm described in Norén et al. (2005). Complete removal of all suspected duplicates is of course overly cautious in the sense that at least one report in each group of suspected duplicates should be retained in the database, but for the purpose of general method evaulation it should have minimal impact on the results. In the future, we intend to implement a more sophisticated approach to account for suspected duplication through report weighting.

All in all, 14,927 cases of three-way disproportional relative reporting rates with $\Omega_{025} > 0$ were highlighted in the database wide screen. Table 2 displays 10 of the drug–drug–ADR triplets with the highest 20 estimated Ω_{025} values in the entire screen. Excluded from the list are 10 drug–drug–ADR triplets that are due to a series of 25 case reports on strabismus together with gentamicin, lidocaine, hyaluronidase, cefazolin and bupivacaine, that fell just below the threshold to be highlighted as suspected duplicates. Further follow up of the three drug–drug–ADR triplets involving cerivastatin and gemfibrozil in Table 2 revealed another potential data quality problem related to a series of some 600 very similar case reports that were originally submitted to a pharmaceutical company by a law firm. While these reports do refer to different patients, they should not be considered as independent pieces of information due to their common origin. Their identification is interesting in its own right. In large ICSR data sets, some data quality issues are unavoidable, and do not negate the value of the proposed method, even though data quality is an important issue in the general use of ICSR systems. That some of the very highest disproportional reporting rates correspond to data quality problems matches experience from pairwise disproportionality analysis. Nevertheless, some of the drug–drug–ADR triplets highlighted in Table 2 are of potential clinical interest. Specifically, the disproportional reporting of med-

Drug 1	Drug 2	ADR	n_{111}	f_{11}	g_{11}	Ω	Ω_{025}
Cerivastatin	Gemfibrozil	Neurological disorder	659	0.31	0.0029	6.63	6.52
Cerivastatin	Gemfibrozil	Heart block	123	0.06	0.0004	6.46	6.19
Celecoxib	Citalopram	Drug abuse	51	0.72	0.0046	5.96	5.53
Cisplatin	Carboplatin	Medication error	118	0.55	0.0084	5.69	5.42
Diphtheria, pertussis, tetanus, poliomyelit	Haemophilus B vaccine	Hypotonic hyporesponsive episode	141	0.033	0.0006	5.49	5.25
Amoxicillin	Cefaclor	Tooth disorder	46	0.41	0.0037	5.68	5.23
Nefazodone	Quetiapine	Medication error	68	0.73	0.0118	5.42	5.06
Metronidazole	Vancomycin	Resistance	21	0.18	0.0006	5.25	4.57
Cerivastatin	Gemfibrozil	Depression	721	0.34	0.0137	4.63	4.52
Donepezil	Rabeprazole	Drug abuse	27	0.26	0.0035	5.00	4.40

Table 2: 10 drug–drug–ADR combinations with among the 20 highest Ω_{025} values in the database wide screen

Drug 1	Drug 2	ADR	n_{111}	f_{11}	g_{11}	Ω	Ω_{025}
Oxycodone	Quetiapine	Suicide attempt	5	0.45	0.13	1.53	0.00
Cisapride	Clarithromycin	Dyspnoea	7	0.10	0.04	1.26	0.00
Diphteria and tetanus toxoids	Haemophilus B vaccine	Face oedema	6	0.04	0.02	1.09	0.00
Furosemide	Amoxicillin	Epidermal necrolysis	4	0.06	0.01	1.74	0.00
Risperidone	Valproic acid	Condition aggravated	17	0.10	0.06	0.77	0.00
BCG vaccine	Interferon alfa-2b	Back pain	3	0.08	0.009	2.05	0.00
Carbamazepine	Thiamine	Fever	3	0.60	0.07	2.05	0.00
Ticlopidine	Acetylsalicylic acid	Death	18	0.06	0.03	0.74	0.00
Haloperidol	Trifluoperazine	Dyskinesia	7	0.23	0.09	1.26	0.00
Phenytoin	Gabapentin	Hypothyroidism	3	0.02	0.002	2.05	0.00

Table 3: 10 drug–drug–ADR combinations with Ω_{025} values just above 0

ication error for concomitant use of cisplatin and carboplatin may indicate a potential patient safety issue. Similarly, the hypotonic, hyporesponsive episodes reported for concomitant administration of two vaccines are in children usually very scary experiences both for the child and its parents. If the interaction is confirmed, some such cases can be avoided by policy changes to vaccine programmes.

The early warning system for pairwise disproportional reporting in the WHO database focuses on relative reporting rates that have just recently crossed a threshold for clinical review Bate et al. (1998). To illustrate what a similar approach to drug–drug interaction screening may generate, we examined the 10 drug–drug–ADR combinations whose Ω_{025} values exceeded 0 with the smallest margin, in our database screen. These are listed in Table 3. Despite the lower relative reporting rates compared to those in Table 2, some of these drug–drug–ADR triplets are also of potential clinical interest. Specifically, ticlopidine and acetylsalicylic acid are anti-platelet drugs that are sometimes co-prescribed for improved potency, and if their co-prescription induces safety problems, this should be accounted for in their clinical management. As for the disproportional reporting of condition aggravated for concomitant use of risperidone and valproic acid, a possible interaction between these two drugs has been discussed in the medical literature (van Wattum 2000).

Some of the examples in Tables 2 and 3 have no obvious pharmacological basis. As such they represent important signals requiring confirmation or explanation. Our aim here is to demonstrate that the proposed measure of disproportionality may generate interesting leads with respect to suspected drug–drug interaction. No clinical assessment has yet been made to exclude spurious associations, confounding by co-medication or underlying disease, and further review of the examples is needed.

4 Discussion

We have introduced a new three-way disproportionality measure for drug–drug interaction, that unlike previously proposed such measures is based on a model with additive risk for the occurrence of ADRs under concomitant use of non interacting drugs. We have showed how an observed-to-expected ratio measure of disproportionality for ADR relative reporting rates, based on this model, can be estimated and used to screen for drug–drug interaction in ICSR data. In addition, we have provided empirical examples of established drug–drug interaction with considerable relative reporting rates in the WHO database that go undetected with other methods such as logistic regression, but can be detected with our approach.

Disproportionality analysis of ICSR data can be seen as a form of case-control study, in which reports on other drugs in the same database are considered as controls for the reporting of the drug of interest. However, by modelling the

additive risk explicitly instead of implementing a logistic regression model, we avoid the potential problems associated with estimating departure from additivity based on a model with essentially multiplicative risk discussed in Skrondal (2003). We use a deliberately rather simple shrinkage for the Ω measure of disproportionality in Section 2.3, much less sophisticated than the complex set of priors for the IC in Norén et al. (2006) and the gamma prior distribution with two components and five fitted parameters used to shrink the EBGM (DuMouchel and Pregibon 2001). The main advantage of this is transparency. Clinical review is a critical step in the knowledge discovery process and reliance on complex statistical methods limits the ability of subject matter experts to interpret and question the relevance of observed disproportional relative reporting rates. For the same reason, we advise against isolated presentation of Ω. Sets of observed and expected relative reporting rates f_{11} (as well as perhaps f_{00}, f_{10} and f_{01}) and g_{11} give subject matter experts a more clear indication why a particular series of case reports has been highlighted for clinical review.

Some of our model assumptions may potentially be violated. While most of these assumptions apply to disproportionality analysis of ICSR data in general, our model formulation makes them explicit. For example the assumption of equal reporting rates r for all drugs, ADRs and combinations thereof in Section 2.2 will sometimes not hold. One can show that (16) is still a valid estimator for the expected relative reporting rate under reporting biases that affect individual drug substances and ADRs separately. However, as in any analysis of ICSRs, the impact of reporting biases that affect specific drug–ADR pairs or drug–drug pairs is more difficult to comment on in general terms. This emphasizes why this and other knowledge discovery methods for ICSRs are tools for hypothesis generation rather than testing. The possibility that an observed disproportional reporting rate is due to complex reporting biases should always be considered in the strengthening and refinement of generated hypotheses. Another violable model assumption is that of a constant risk of the reference set of ADRs A' for all combinations of D_1 and D_2 in (11) of Section 2.1. In reality, interaction between D_1 and D_2 may increase the overall risk for ADRs other than A. If so, Ω will under-estimate the disproportionality of the observed relative reporting rate — much like the phenomenon referred to as masking in pairwise disproportionality analysis of ICSR data sets, where excessive reporting on a specific ADR for a certain drug masks less extreme disproportional reporting of the same ADR given other drugs (Evans 2004). As stated above, this can be remedied by restricting A' to a more limited set of ADRs.

The discovery in Section 3.2 of a cluster of ICSRs provided by the same law firm illustrates the importance of further analysis of observed disproportional reporting rates. While suspicions based on ICSRs remain tentative even after clinical review, clusters of ICSRs with a reasonable spread in space and time, cleaned from case report duplication and other reporting biases, provide stronger indication. Possible confounders should also, as far as possible, be ruled out as alternative explanations. The quality and amount of information on highlighted ICSRs is very important in the clinical review. Suspected

ADR incidents are often originally described in pieces of free text, only later encoded in terms of standard ADR terminologies. If this conversion is not satisfactory, it may distort any subsequent analysis. The two references to Drug abuse in Table 2, may be examples of this. The term Drug abuse has diverse possible interpretations, and careful review of the original reports listing drug abuse for celecoxib together with citalopram indicates that they actually refer to instances of medication error, where the two drugs have not been taken together but one has mistakenly been dispensed instead of the other, on account of their similar commercial names (Celebrex and Celexa). While not a drug–drug interaction per se, we consider it beneficial that our method highlights this interesting association between two drugs and one ADR.

The work presented here shall need to be complemented in the future by applied method development with the aim of presenting a routine framework for drug–drug interaction surveillance in the WHO database. Important challenges include the definition of effective triage strategies to focus efforts in drug–drug interaction surveillance on the most important issues for follow-up, similar to those developed for pairwise drug–ADR disproportionality analysis by Ståhl et al. (2004). Clearly, strategies to incorporate increased pharmacological knowledge such as that related to pharmacogenetics may also improve the potential for effective drug–drug interaction detection (Strandell et al. 2005). A framework for hypothesis strengthening and refinement related to highlighted case series must also be developed and implemented.

References

Almenoff, J. S., DuMouchel, W., Kindman, L. A., Yang, X. and Fram, D.: 2003, Disproportionality analysis using empirical Bayes data mining: a tool for the evaluation of drug interactions in the post-marketing setting, *Pharmacoepidemiology and Drug Safety* **12**(6), 517–521.

Bate, A., Lindquist, M., Edwards, I. R., Olsson, S., Orre, R., Lansner, A. and De Freitas, R. M.: 1998, A Bayesian neural network method for adverse drug reaction signal generation, *European Journal of Clinical Pharmacology* **54**, 315–321.

DuMouchel, W.: 1999, Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system, *American Statistician* **53**, 177–202.

DuMouchel, W. and Pregibon, D.: 2001, Empirical Bayes screening for multi-item associations, *KDD '01: Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 67–76.

Edwards, I. R., Wiholm, B.-E., Lindquist, M. and Napke, E.: 1990, Quality criteria for early signals of possible adverse drug reactions, *Lancet* **336**(8708), 156–158.

Evans, S. J. W.: 2004, Statistics: analysis and presentation of safety data, *in* J. Talbott and P. Waller (eds), *Stephens' detection of new adverse drug reactions*, Wiley.

Lindquist, M., Ståhl, M., Bate, A., Edwards, I. R. and Meyboom, R. H. B.: 2000, A retrospective evaluation of a data mining approach to aid finding new adverse drug reaction signals in the WHO international database, *Drug Safety* **23**(6), 533–542.

Norén, G. N., Bate, A., Orre, R. and Edwards, I. R.: 2006, Extending the methods used to screen the WHO drug safety database towards analysis of complex associations and improved accuracy for rare events, *Statistics in Medicine* **25**(21), 3740–3757.

Norén, G. N., Orre, R. and Bate, A.: 2005, A hit-miss model for duplicate detection in the WHO drug safety database, *KDD '05: Proceeding of the eleventh ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM Press, New York, NY, USA, pp. 459–468.

Olsson, S.: 1998, The role of the WHO programe on international drug monitoring in coordinating worldwide drug safety efforts, *Drug Safety* **19**(1), 1–10.

Pirmohamed, M. and Orme, M.: 1998, Drug interactions of clinical importance, *in* D. Davies, R. Ferner and H. de Glanville (eds), *Davies's Textbook of Adverse Drug Reactions*, 5 edn, Chapman & Hall Medical, pp. 888–912.

Rawlins, M. D.: 1988, Spontaneous reporting of adverse drug reactions. II: Uses, *British Journal of Clinical Pharmacology* **1**(26), 7–11.

Rothman, K. J., Greenland, S. and Walker, A. M.: 1980, Concepts of interaction, *American Journal of Epidemiology* **112**(4), 467–470.

Skrondal, A.: 2003, Interaction as departure from additivity in case-control studies: a cautionary note, *American Journal of Epidemiology* **158**(3), 251–258.

Stephens, M. D. B.: 2004, Ethical issues in drug safety, *in* J. Talbott and P. Waller (eds), *Stephens' detection of new adverse drug reactions*, Wiley.

Strandell, J., Kiuru, A. and Lindquist, M.: 2005, Linking spontaneous reports with pharmacogenetics, *Basic and Clinical Pharmacology and Toxicology* **97**, 106.

Ståhl, M., Lindquist, M., Edwards, I. R. and Brown, E. G.: 2004, Introducing triage logic as a new strategy for the detection of signals in the WHO drug monitoring database, *Drug Safety* **13**(6), 355–363.

van Puijenbroek, E. P., Egberts, A. C., Heerdink, E. R. and Leufkens, H. G. M.: 2000, Detecting drug-drug interactions using a database for spontaneous adverse drug reactions: an example with diuretics and non-steroidal anti-inflammatory drugs, *European Journal of Clinical Pharmacology* **56**, 733–738.

van Puijenbroek, E. P., Egberts, A. C., Meyboom, R. H. B. and Leufkens, H. G. M.: 1999, 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* **47**, 689–693.

van Wattum, P. J.: 2000, Letter to the editor: Valproic acid and risperidone, *Journal of the American Academy of Child & Adolescent Psychiatry* **40**(8), 866–867.

Yang, X. and Fram, D.: 2004, Using disproportional analysis as a tool to explore severe drug-drug interactions in AERS database, *Pharmacoepidemiology and Drug Safety* **13**(1), S247.