

# Safety Related Drug-Labeling Changes Findings from Two Data Mining Algorithms

Manfred Hauben<sup>1,2,3</sup> and Lester Reich<sup>1</sup>

1 Risk Management Strategy, Pfizer Inc, New York, New York, USA

2 Department of Medicine, New York University School of Medicine, New York, New York, USA

3 Departments of Pharmacology and Community and Preventive Medicine, New York Medical College, Valhalla, New York, USA

## Abstract

**Introduction:** With increasing volumes of postmarketing safety surveillance data, data mining algorithms (DMAs) have been developed to search large spontaneous reporting system (SRS) databases for disproportional statistical dependencies between drugs and events. A crucial question is the proper deployment of such techniques within the universe of methods historically used for signal detection. One question of interest is comparative performance of algorithms based on simple forms of disproportionality analysis versus those incorporating Bayesian modelling. A potential benefit of Bayesian methods is a reduced volume of signals, including false-positive signals.

**Objective:** To compare performance of two well described DMAs (proportional reporting ratios [PRRs] and an empirical Bayesian algorithm known as multi-item gamma Poisson shrinker [MGPS]) using commonly recommended thresholds on a diverse data set of adverse events that triggered drug labelling changes.

**Methods:** PRRs and MGPS were retrospectively applied to a diverse sample of drug-event combinations (DECs) identified on a government Internet site for a 7-month period. Metrics for this comparative analysis included the number and proportion of these DECs that generated signals of disproportionate reporting with PRRs, MGPS, both or neither method, differential timing of signal generation between the two methods, and clinical nature of events that generated signals with only one, both or neither method.

**Results:** There were 136 relevant DECs that triggered safety-related labelling changes for 39 drugs during a 7-month period. PRRs generated a signal of disproportionate reporting with almost twice as many DECs as MGPS (77 vs 40). No DECs were flagged by MGPS only. PRRs highlighted DECs in advance of MGPS (1–15 years) and a label change (1–30 years). For 59 DECs, there was no

signal with either DMA. DEC's generating signals of disproportionate reporting with only PRRs were both medically serious and non-serious.

**Discussion/conclusion:** In most instances in which a DEC generated a signal of disproportionate reporting with both DMAs (almost twice as many with PRRs), the signal was generated using PRRs in advance of MGPS. No medically important events were signalled only by MGPS. It is likely that the incremental utility of DMAs are highly situation-dependent. It is clear, however, that the volume of signals generated by itself is an inadequate criterion for comparison and that clinical nature of signalled events and differential timing of signals needs to be considered. Accepting commonly recommended threshold criteria for DMAs examined in this study as universal benchmarks for signal detection is not justified.

## Introduction

The principal concern of pharmacovigilance is the timely discovery of adverse events (AEs) that are novel in terms of their clinical nature, severity, and/or frequency as early as possible after marketing with minimum patient exposure. With the ever-increasing volume of postmarketing safety surveillance data, computer-assisted signal detection algorithms, also known as data mining algorithms (DMAs), have been developed that can search extremely large spontaneous reporting system (SRS) databases for disproportional statistical dependencies between drugs and events relative to the generality of the database or in excess of what would be expected if the drug and event were independently distributed in the database.<sup>[1-4]</sup> If there is sufficient correlation between the observed statistical dependencies and causal relationships, DMAs could significantly improve our ability to detect early 'signals' of AEs. If the majority of such statistical dependencies are not associated with causal associations and/or do not predate signals generated by clinical observations, then the potential added utility of such techniques is limited. DMAs include simple disproportionality analysis such as proportional reporting ratios (PRRs)<sup>[4]</sup> and reporting odds ratios (RORs)<sup>[5,6]</sup> and algorithms that utilise Bayesian in-

ference to adjust for data variance such as the multi-item gamma Poisson shrinker (MGPS)<sup>[2,3]</sup> and the Bayesian Confidence Propagation Neural Network (BCPNN).<sup>[1]</sup>

Although the precise operational details of each method differ, they all use the background frequency of drugs and events in the database as an internal control, instead of turning to external data sets for exposure data. Each method calculates an observed to expected ratio for each drug-event combination (DEC) based on this internal control. The Bayesian methods are designed to down weigh the observed to expected ratio scores, especially those associations based on small numbers of reports, since such estimates are presumably less statistically stable. This is accomplished by the process of Bayesian inference itself (e.g. shrinkage to the null) as well as by additional data transformations, threshold selection, and/or covariate adjustments. While the Bayesian approach may improve the signal to noise ratio, it may be associated with some decreased capacity to detect signals earlier when commonly cited thresholds are used.

Both false-positive and false-negative signalling have adverse impacts on drug safety surveillance and public health. Spurious associations (false-positives) are considered to be particularly common with spontaneous reporting databases that are populated

with voluntarily reported data that are uncontrolled, incomplete with respect to important clinical and demographic variables, and susceptible to numerous reporting biases and confounders.<sup>[7,8]</sup> Missed or delayed signalling of important medical events can have profoundly deleterious consequences on public health. However, the adverse public health impact of false-positive findings should not be underestimated, since they can divert precious pharmacovigilance resources and further erode confidence in the signalling process.

An objective of developing certain DMAs incorporating data transformations, covariate adjustment, and Bayesian shrinkage is elimination of false-positive findings or spurious associations related to confounding or the statistical instabilities associated with low reporting frequencies. Given the degree of confounding, both measured and unmeasured, in SRS databases, and additional complexities such as effect modification, it is unlikely that statistical adjustments or modelling can completely eliminate spurious associations and could in fact produce spurious associations as well.<sup>[9]</sup>

In fact, a consistent finding in the published literature is the significantly lower volume of 'signals' generated by Bayesian versus simple algorithms. The volume of signals generated in some settings with PRRs (and with one of the Bayesian methods that did not routinely include covariate adjustments as well) has required the application of additional triage criteria for signal selection.<sup>[10]</sup> This could represent a significant drawback of simple disproportionality analysis relative to Bayesian methods with covariate adjustment. This conclusion may be mitigated for the following reasons: (i) the clinical nature of the AEs that are 'filtered out' by Bayesian methods using commonly recommended thresholds is still not well defined, may include medically important causal associations, and may be highly situation dependent; (ii) these methods are not suitable for stand-alone signal detection and the

impact of differential signal volumes and performance characteristics are probably substantially mitigated by the concurrent application of traditional, clinical rule-based criteria for identifying signals; and (iii) there is limited or no published data on the differential timing of signal generation in the simple versus Bayesian methods with DECAs that are highlighted by both. Therefore, the differential volume of signals generated by the different methods is not by itself an adequate criterion for judging their relative merits.

To date, drawing conclusions from the published literature is complicated by various limitations and discordant results related to the retrospective nature of most analyses, non-systematic samples of drugs and events, lack of standardised data mining practices, inherent differences between algorithms or models, lack of criterion standards for adjudicating causality and expectedness, and/or variable study design and database and dictionary architecture/environments.<sup>[1-6,11-13]</sup>

Given the above considerations, the comparative assessment of these methods requires careful consideration of not only differential signal volume, but also differential timing and the clinical nature of the events that may be filtered out, or susceptible to delayed recognition, by various models or statistical data adjustments. To put it another way, it is important to understand if and how many important causal associations are concurrently missed or susceptible to delayed recognition, by diminishing the number of false-positive signals through the statistical procedures and thresholds incorporated into Bayesian DMAs.

The objective of this analysis is to understand the clinical nature of the events that may be filtered out by DMAs that include statistical procedures and adjustments designed to minimise false-positive findings and the associated potential clinical impacts by comparing one form of simple disproportionality analysis (PRRs) to a well described empiri-

cal Bayesian algorithm (MGPS) on a diverse set of DEC's that triggered safety-related labelling changes. It is hoped that this will contribute to the collective knowledge of the function of DMAs in a variety of settings and with a variety of drugs and events and stimulate further discussion and research on the application of DMAs. This knowledge may improve the ability to perform comparative assessments of these methods, promote their optimum application, and minimise the potential for their misapplication and misuse.

## Methods

### Drug-Event Combination/Labelling Change Selection

MedWatch is the US FDA Safety Information and Adverse Event Reporting Program. It allows for the reporting of safety problems associated with drugs and medical devices into the FDA Adverse Event Reporting System (AERS). Safety information is disseminated to the healthcare community and the public-at-large via safety alerts, recalls, withdrawals, and labelling changes. Safety-related labelling changes are posted (on a monthly basis since 1997) on the MedWatch website in the section 'Summary of Safety-Related Labeling Changes Approved by FDA Center for Drug Evaluation and Research (CDER)'.<sup>[14]</sup>

For the current analysis, the safety-related drug labelling changes that were posted on the MedWatch website between 01 June 2001 and 31 December 2001 were manually reviewed to identify a suitable sample of AEs that had been added to the 'Adverse Reactions' section of any product label during this time period. The time period was chosen because safety-related labelling changes were colour-coded commencing with the June 2001 posting, which facilitated accurate data extraction. The analysis was limited to simple DEC's not previously

listed in 'Adverse Reactions' sections that were not derived from clinical trials. Drug-drug interactions were not examined because they are ordinarily not contained in the 'Adverse Reactions' section and would have required separate and more computationally intensive data mining. For each specific AE, the verbatim term for data mining was used. If the verbatim term did not correspond to a Medical Dictionary for Medical Regulatory Activities (MedDRA) Preferred Term, all MedDRA Preferred Terms that were considered clinically equivalent or closely related to the verbatim term were used. Each author performed the data mining analysis. Discrepancies in the results were identified and adjudicated between the authors.

### Adverse Event Data Set

The AE data set for this analysis consisted of an extract of the FDA AERS database.<sup>[14]</sup> This is a computerised information database for post-approval safety surveillance. It functions as an early warning system for adverse drug reactions not detected during pre-approval testing. It contains AE reports with approved drugs and therapeutic biological products submitted in accordance with mandatory reporting obligations by pharmaceutical companies and voluntarily by healthcare professionals and consumers. Adverse event reports are reviewed and coded for data entry in accordance with the standardised terminology of MedDRA. Quarterly extracts are available through the National Technical Information Service (NTIS). These quarterly updates were subjected to extensive cleaning to correct for report duplication and redundant drug nomenclature prior to data mining. The data extract used for the current analysis included data in AERS from 1968 through the second quarter of 2002.<sup>[15]</sup> The temporal resolution of the data mining was 1 year.

## Data Mining Process

The two data mining algorithms chosen for this analysis were PRRs<sup>[4]</sup> and the empirical Bayesian algorithm MGPS (Lincoln Technologies, Wellesley Hills, Massachusetts, USA).<sup>[3]</sup> Data mining was performed individually on the MedDRA Preferred Terms synonymous or clinically compatible with the AE term added to the product label.

The PRR is a simple metric relating the proportional representation of an event of interest with a drug of interest compared with the proportional representation of that event among all other drugs in the database (figure 1). For this analysis a PRR >2 with an associated  $\chi^2 >4$  (with Yates correction) was considered a 'signal' of disproportionate reporting, which has been frequently used in published studies of data mining.<sup>[4]</sup>

The theoretical basis of MGPS has been described in detail elsewhere<sup>[2,3]</sup> but briefly is as follows. Expected counts for item sets (i.e. DEC)s are related to the product of the marginal probabilities of each item (drug and event). The observed to expected (O/E) ratio is initially calculated as a crude disproportionality metric. Since the same ratio could be obtained from cell counts (frequencies) of markedly different sizes (O/E ratios based on smaller cell counts being considered more variable or imprecise) further modelling using maximum likelihood estimation and Bayesian inference are used to adjust the crude O/E ratios based on the respective cell counts. Each cell is considered to represent a Poisson process in which the Poisson parameter distribution is a mixture of two gamma distributions. The prior

probability distribution of the gamma parameters are obtained by applying an iterative maximum likelihood algorithm to a negative binomial mixture likelihood. Posterior estimates of the gamma parameters are obtained by updating the prior with the individual cell counts via Bayes theorem.

By using logarithmic transformations or taking the lower 5% cut-off of the posterior distribution (EB05), an expectation value that adjusts for the variability by down weighting or 'shrinking' the parameters associated with low cell counts is obtained. These metrics are known as the empirical Bayes geometric mean (EBGM) and the EB05. An EB05 of 8 is therefore interpreted to mean that reports of the particular DEC occur in the database eight times more frequently than would be expected if drug and event were independently distributed in the database. The signal metric used for a threshold in the current analysis was the lower 5% cut-off of the distribution of the empirical Bayes geometric mean  $\geq 2$  (EB05  $\geq 2$ ), which has been frequently recommended in published studies of data mining.<sup>[3]</sup> A variety of data mining options and parameters are available in MGPS including basic covariate adjustment (stratification by age, gender, and year of report). Stratification tends to reduce spurious associations due to confounding and markedly decreases the volume of disproportionalities.<sup>[2,3]</sup>

For the present analysis, the data mining was performed on suspect drug-AE pairs, using stratification by age, gender and FDA year of reporting with cumulative subsetting by year (for EB05 calculations).

	Drug of interest	All other drugs
Reaction(s) of interest	A	B
All other reactions	C	D

$$\text{PRR} = \frac{A/(A + C)}{B/(B + D)}$$

Fig. 1. Calculation of the proportional reporting ratio (PRR).

## Comparative Analysis

Metrics for the comparative analysis included the number and proportion of DEC)s that generated signals of disproportionate reporting with PRRs, MGPS, both or neither method, differential timing of signal generation between the two methods, and

the clinical nature of events that generated signals with only one, both or neither method.

## Results

The manual review of the 'Summary of Safety-Related Labeling Changes Approved by FDA Center for Drug Evaluation and Research (CDER)' for the period 01 June 2001 to 31 December 2001 identified 136 DEC's that triggered safety-related labelling changes involving 39 drugs. PRRs generated a signal of disproportionate reporting for almost twice as many of these DEC's as MGPS (77 vs 40). There were no DEC's that were flagged by MGPS only. Table I summarises the number and proportion of DEC's highlighted by each method.

Differences were observed in the timing of signals between the two methods. PRRs always highlighted DEC's in advance of MGPS (1–15 years). Another notable finding was that both PRRs and MGPS usually flagged DEC's in advance of the label change (1–30 years) although PRR was more likely to highlight an association two or more years in advance of the label change than MGPS. PRRs were three times as likely to highlight an association over 10 years in advance of a labelling change. Table II displays the differential timing of signals among those DEC's that were highlighted by both.

There were 69 cases where only PRR 'signalled' or PRR 'signalled' prior to MGPS. Forty-five of the 69 cases 'signalled' with three or more reports. Table

**Table I.** Safety-related drug labelling changes: number and proportion of drug-event combinations reported by proportional reporting ratios (PRRs) and multi-item gamma Poisson shrinker (MGPS) [n = 136]

	Signalled with MGPS	Not signalled with MGPS	Total
Signalled with PRRs	40 (29.4%)	37 (27.2%)	77 (56.6%)
Not signalled with PRRs	0 (-)	59 (43.4%)	59 (43.4%)
<b>Total</b>	<b>40 (29.4%)</b>	<b>96 (70.6%)</b>	<b>136</b>

**Table II.** Drug-event combinations (DECs) signalled by both proportional reporting ratios (PRRs) and multi-item gamma Poisson shrinker (MGPS) [n = 40]<sup>a</sup>: differential timing

	Signalled with PRRs first (n = 32)
By 1 year	9
By 2–5 years	18
By 6–10 years	4
By >10 years	1
<b>Total</b>	<b>32</b>

<sup>a</sup> Eight DEC's signalled in same year for PRR and MGPS.

III shows the distribution by number of cases required to signal for PRR.

There were many medically important events that generated signals only with PRRs and/or with PRRs in advance of MGPS (table IV).

Non-serious events highlighted by PRRs only included flatulence, arthralgia, malaise, gynaecomastia, oedema, bruising, abdominal pain, confusion, sicca syndrome, somnolence, misdirected eyelashes, pruritus, alopecia, and eosinophilia. Non serious events highlighted by PRRs in advance of MGPS include oedema/peripheral oedema, myoclonia, hypertension, glucose intolerance, gastroenteritis, fat redistribution/accumulation, ataxia, taste perversion, and tachycardia.

Fifty-nine DEC's were not highlighted by PRRs and 96 were not highlighted by MGPS. There were a wide variety of events for the 59 DEC's, both medically serious and non serious, that did not generate signals of disproportionate reporting with either method (e.g. anaphylaxis, hepatic failure, congestive heart failure, toxic epidermal necrolysis, insomnia, anxiety). For seven DEC's (mostly non-serious), there were no cases in AERS.

## Discussion

While the findings of the current analysis are consistent with previous investigations demonstrating the lower signal volume generated by MGPS compared with PRRs,<sup>[10]</sup> it underscores the fact that such a diminished signal volume may possibly have

deleterious consequences in the form of missed or delayed signalling.

PRRs generated signals of disproportionate reporting with almost twice as many DEC associated with safety-related labelling changes as MGPS. In most instances in which a DEC generated a signal of disproportionate reporting by both methods, a signal was generated using PRRs in advance of MGPS. There were important medical events that were signalled only by PRRs but no medically important DECs that were signalled only by MGPS. It notably demonstrates that certain procedures and statistical adjustments may filter out, either absolutely or relatively in terms of timing, medically important associations when commonly cited thresholds are used. Subsequent analysis of additional data from AERS resulted in the observation of similar performance gradients.<sup>[16,17]</sup> Therefore, all three factors, signal volume, differential timing of signals, and the clinical nature of the signalled versus non-signalled events need to be carefully considered when comparing methods.

For 59 DECs, there was no signal with either DMA. There are numerous mechanisms that may explain the failure of a DMA to generate a signal. The most basic relates to the very nature of the disproportionality analysis that underlies the commonly used DMAs. DECs that are 'over-represented' in a database of finite size (at a given time point) must be accompanied by other DECs being under-represented. Therefore, if a drug is strongly associated with one or more AEs, the latter features of the drug's safety profile may 'crowd out' signals of other AEs that may be causally related. Also, if the background prevalence of the drug or the event in the database is high, this could result in failure to signal, since 'expected' counts are derived from the

marginal probabilities of drug and event in the database. This, in combination with the numerous variables involving reporting behaviour, provides a conceptual framework for understanding missed or delayed signalling.

There are significant limitations to this analysis. Quantitatively, this data set of DECs represents a tiny fraction of DECs in the AERS database, did not include AEs contained in the 'Warnings, Precautions, and Contraindications' section of drug labelling, and therefore may not be fully representative, although it is diverse in terms of drugs and events. It should be considered a case series with the focused objective of describing the clinical nature of the AEs that may be filtered out or subject to delayed recognition by statistical adjustments/modelling/thresholds employed by some DMAs. It is not a recommendation for or against either DMA. Qualitatively, a data set of AEs that triggered a postmarketing safety-related labelling change should not be considered an ideal criterion standard for adjudicating causality since an AE may be added to labelling for reasons unrelated to causality (e.g. regulatory request, various legal considerations). However, a limited clinical review of the published literature on a subset of DECs confirmed that there were medically important associations included in the safety related labelling changes, many of which were identified only by PRRs or earlier by PRRs than MGPS, using commonly cited thresholds.

This analysis is not designed to derive estimates of the rate or volume of false positive findings with either method. Although it might be postulated that some of the DECs signalled only by PRRs represent false positive signalling, important findings persisted when the analysis was limited to events that were signalled by both methods. Still, the analysis is not a

**Table III.** Frequency distribution of number of cases required to 'signal' for proportional reporting ratios (PRR)

No. of cases required to 'signal' for PRR	1	2	3	4	5	6-10	11-20	21-4
Case count	12	12	10	6	8	9	9	3

**Table IV.** Medically significant events<sup>a</sup> highlighted only by proportional reporting ratios (PRRs) or highlighted by PRRs in advance of multi-item gamma Poisson shrinker (MGPS)

---

**Signalled only with PRRs**

Aggressive behaviour  
 Angina pectoris  
 Aplastic anaemia  
 Anaphylactoid reaction  
 Angioedema  
 Diabetic coma/ketoacidosis  
 Hepatitis  
 Increased International Normalized Ratio/prothrombin time  
 Interstitial pneumonitis  
 Leucocytoclastic vasculitis/vasculitis  
 Neuroleptic malignant syndrome  
 Pancreatitis  
 Psychosis  
 Renal failure  
 Seizure  
 Severe hypoglycaemia  
 Stevens-Johnson syndrome  
 Thrombocytopenia  
 Ventricular tachycardia/fibrillation

**Signalled with PRRs prior to MGPS**

Agranulocytosis  
 Anaphylaxis/anaphylactic reaction  
 Decreased consciousness  
 Encephalopathy  
 Hallucination  
 Hepatic failure/liver transplant/hepatic necrosis/hepatitis/abnormal hepatic function/liver function tests abnormal  
 Henoch Shoenlein purpura  
 Intestinal obstruction  
 Myopathy  
 Pulmonary toxicity  
 Severe hyperglycaemia  
 Tendon rupture  
 Torsade de pointes/ventricular tachycardia/cardiac arrest/QT-related events  
 Toxic epidermal necrolysis  
 Vasculitic syndrome

---

a These events may have been signalled only with PRRs for certain drugs, signalled with PRRs prior to MGPS for other drugs, and/or not signalled with either data mining algorithm in certain instances. Serious adverse events signalled at the same time are not included in this table.

---

systematic assessment of the comparative performance of the two methods nor an endorsement of either method.

We did not assess the relative contributions of the thresholds used versus the Bayesian process itself (e.g. shrinkage to the null), additional data transformations, statistical procedures, and/or extensive covariate adjustments that are features of MGPS, to the observed results. We used only the most commonly cited of the many possible combinations of PRR and MGPS threshold specifications. Users of PRRs have employed various thresholds based on the value of the PRR, the associated  $\chi^2$  value, and/or a minimum number of reports. Similarly in MGPS, one can use the EBGM, which generates more signals and therefore might perform closer to PRRs than EB<sub>05</sub>, as well as lower EB<sub>05</sub> thresholds. There were a significant number of DEC's highlighted by PRRs >10; therefore, significantly higher thresholds may still provide comparable performance possibly with a reduced signal volume. Therefore, it would be useful to have systematic comparisons of various combinations of threshold specifications for each method because these commonly cited thresholds (EB<sub>05</sub> ≥ 2 and PRR > 2 with an associated  $\chi^2$  > 4) may not be applicable to every user in every situation. Additionally, there is no universal definition of a signal.

Data mining is a very active field. There are numerous data mining strategies and configurations and as DMAs continue to evolve they will include enhancements that might result in improved performance, such as the ability to combine clinically equivalent or compatible AE terms into a single variable for purposes of data mining. This may be particularly relevant given the granularity of MedDRA. Studying the performance of these techniques should be ongoing with such technical enhancements. There has been less research on the clinical judgement and heuristics that have tradition-



ally been used to identify signals and this should not be neglected.

Finally, the identification of a signal of disproportionate reporting, even of causally related DECs does not ensure that the causal nature of the association would have been recognised at the time the initial signal was evaluated. We also could not determine the time at which the DEC was prospectively identified as a signal. Safety-related labelling changes are often the end result of a process of signal detection and evaluation that take variable amounts of time. Therefore, there remains a significant degree of residual uncertainty with respect to the conclusions that can be drawn with respect to chronology.

## Conclusion

Developing an optimal pharmacovigilance strategy should be individualised based on the nature of the drugs, events, and database environment being considered. Performance gradients between traditional rule-based methods, traditional methods plus simple disproportionality analysis (e.g. PRRs) or traditional methods plus Bayesian methods remain unclear. The ideal point on the aforementioned performance gradient between simple disproportionality analysis and available empirical Bayesian methods for decision making will be highly situation dependent. Data mining algorithms, if used, should only be used as supplements to, and not substitutes for standard signalling strategies. The observed performance differentials are likely to be significantly mitigated when these methods are used as one element of a comprehensive pharmacovigilance programme. Therefore, in assessing the potential added value of DMAs and/or selecting a specific algorithm, potential users of DMAs should consider numerous factors such as the rigour of their standard signalling criteria, the nature and numbers of drugs and AE reports to be screened, relative tolerance of false-positive and false-negative findings, timing in

the product life cycle, and resource constraints. It is quite likely that the incremental utility of DMAs may be higher for health authorities, who have statutory obligations for monitoring the safety of all licensed drugs, than for an individual pharmaceutical company whose surveillance responsibilities are more circumscribed. Data mining algorithms are promising tools, however, any institution contemplating the use of DMAs should be aware of the multiple elements that should enter into a comparative assessment, including differentials in the clinical nature and timing of signalled and non-signalled DECs. Further research should examine a variety of threshold criteria for each method being examined in combination with rigorous clinical criteria for identifying signals.

## Acknowledgements

We would like to express our sincere thanks to our intern, Miss Stephanie Chung, for her help in preparation for this paper.

No sources of funding were used to assist in conducting this study. The authors have no conflicts of interest that are directly relevant to the content of this study.

## References

1. Bate A, Lindquist M, Edwards I, et al. A Bayesian neural network method for adverse drug reaction signal generation. *Eur J Clin Pharmacol* 1998; 54: 315-21
2. DuMouchel W. Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. *Am Stat* 1999; 53 (3): 170-90
3. Szarfman A, Machado SG, O'Neill R. Use of screening algorithms and computer systems to efficiently signal higher-than expected combinations of drugs and events in the US FDA's spontaneous reports database. *Drug Saf* 2002; 25 (6): 381-92
4. Evans SJW, Waller P, Davis S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiol Drug Saf* 2001; 10: 483-6
5. Van Puijenbroek E, Diemont W, van Groothest K. Application of quantitative signal detection in the Dutch spontaneous reporting system for adverse drug reactions. *Drug Saf* 2003; 26 (5): 293-301
6. Van der Heijden P, van Puijenbroek E, van Buuren S, et al. On the assessment of adverse drug reactions from spontaneous reporting systems: the influence of under-reporting on odds ratios. *Stat Med* 2002; 21: 2027-44

7. Kennedy D, Goldman S, Lillie R. Spontaneous reporting systems in the United States. In: Strom B, editor. *Pharmacoepidemiology*. 3rd ed. New York: Wiley, 2000: 151-74
8. Wiholm B, Olsson S, Moore N, et al. Spontaneous reporting systems outside the US. In: Strom B, editor. *Pharmacoepidemiology*. 3rd ed. New York: Wiley, 2000: 175-92
9. Reintjes R, de Boer A, van Pelt W. Simpson's paradox: an example from hospital epidemiology. *Epidemiology* 2000; 11 (1): 81-3
10. Hauben M. A brief primer on automated signal detection. *Ann Pharmacother* 2002; 37 (7/8): 1117-26
11. Lindquist M, Stahl M, Bate A. A retrospective evaluation of a data mining approach to aid finding new adverse drug reaction signals in the WHO international database. *Drug Saf* 2000; 23 (6): 533-42
12. Gould AL. Practical pharmacovigilance analysis strategies. *Pharmacoepidemiol Drug Saf* 2003; 12 (7): 559-74
13. Van Puijenbroek E, Bate A, Leufkens H, et al. A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions. *Pharmacoepidemiol Drug Saf* 2002; 11 (1): 3-10
14. Medwatch: the FDA safety information and adverse drug event reporting program [online]. Available from URL: <http://www.fda.gov/medwatch/index.html> [Accessed 2004 Jul 12]
15. Center for Drug Evaluation and Research. Drug information: Adverse Event Reporting System (AERS) [online]. Available from URL: <http://www.fda.gov/cder/aers/default.htm> [Accessed 2004 Jul 12]
16. Hauben M. Trimethoprim-induced hyperkalemia: lessons in data mining. *Br J Clin Pharmacol*. DOI:10.1111/j.1365-2125.2004.02153.x [online]. Available from URL: <http://www.blackwell-synergy.com/> [Accessed 2004 Jul 12]
17. Hauben M, Reich L. Drug-induced pancreatitis: lessons in data mining. *Br J Clin Pharmacol*. In press

---

Correspondence and offprints: Dr *Lester Reich*, Risk Management Strategy, Pfizer Inc., 235 E. 42nd St, NY 10017, USA.

E-mail: [Lester.Reich@Pfizer.com](mailto:Lester.Reich@Pfizer.com)