CURRENT OPINION

Drug Safety 2005; 28 (10): 835-842
0114-5916/05/0010-0835/$34.95/0

Data Mining in Pharmacovigilance
The Need for a Balanced Perspective

Manfred Hauben,[1,2,3] *Vaishali Patadia*,[4] *Charles Gerrits*,[5] *Louisa Walsh*[6] and *Lester Reich*[1]

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, Valhalla, New York, USA
4 Global Drug Safety, Amylin Pharmaceuticals, San Diego, California, USA
5 Department of Pharmacoepidemiology and Outcomes Research, Takeda Global Research and Development, Lincolnshire, Illinois, USA
6 Clinical Drug Safety, AstraZeneca LP, Wilmington, Delaware, USA

Abstract

Data mining is receiving considerable attention as a tool for pharmacovigilance and is generating many perspectives on its uses. This paper presents four concepts that have appeared in various professional venues and represent potential sources of misunderstanding and/or entail extended discussions: (i) data mining algorithms are unvalidated; (ii) data mining algorithms allow data miners to objectively screen spontaneous report data; (iii) mathematically more complex Bayesian algorithms are superior to frequentist algorithms; and (iv) data mining algorithms are not just for hypothesis generation. Key points for a balanced perspective are that: (i) validation exercises have been done but lack a gold standard for comparison and are complicated by numerous nuances and pitfalls in the deployment of data mining algorithms. Their performance is likely to be highly situation dependent; (ii) the subjective nature of data mining is often underappreciated; (iii) simpler data mining models can be supplemented with 'clinical shrinkage', preserving sensitivity; and (iv) applications of data mining beyond hypothesis generation are risky, given the limitations of the data. These extended applications tend to 'creep', not pounce, into the public domain, leading to potential overconfidence in their results. Most importantly, in the enthusiasm generated by the promise of data mining tools, users must keep in mind the limitations of the data and the importance of clinical judgment and context, regardless of statistical arithmetic. In conclusion, we agree that contemporary data mining algorithms are promising additions to the pharmacovigilance toolkit, but the level of verification required should be commensurate with the nature and extent of the claimed applications.

"Data mining has promise, but there are many difficulties associated with it. It is not to be entered into lightly or in ignorance of the obstacles."[1]

The principle concern of pharmacovigilance is the timely discovery of adverse drug reactions 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 spontaneous reporting system data, considerable research is

being devoted to computer-assisted signal detection algorithms (also known as data mining algorithms).[2-9] Contemporary data mining algorithms can screen extremely large spontaneous reporting system databases for statistical dependencies between drugs and events in excess of what would be expected if the drug and event were independently distributed in the database (so-called 'disproportionality analysis'). If there is sufficient correlation between these observed statistical dependencies and demonstrable causal relationships, data mining algorithms could significantly improve our ability to detect signals of novel adverse events early in the postmarketing period.

Despite the significant research published on these methods in the last few years, misconceptions and pitfalls remain when these methods are investigated or used in daily pharmacovigilance practice. The goal of this commentary is to point out some ideas and descriptions that might lead to misunderstanding, particularly by newcomers to the field, and to present alternative or extended viewpoints. Some of these ideas and descriptions appear in previously published articles, others are the result of information communicated at professional societies and meetings that are as yet unpublished. When appearing in the published literature they may be subsidiary to the main body of the text. Regardless of the source, our collective experience in the field of data mining suggests that potentially confusing statements that are initially innocuous when occurring in isolation may become impediments to formulating a balanced perspective when they are repeated and gradually permeate into a wider audience. We hope that by repairing misconceptions, clarifying fuzzy statements and extending discussions we will promote better communication and evaluation practices, interpretation of the published literature and deployment of these tools in naturalistic pharmacovigilance settings. Because of space limitations, we have focused our exposition on four discussion points followed by closing comments. Our first discussion point extends discussions presented in the published literature, while the remaining three address specific statements and concepts that are potential sources of misunderstanding.

1. Data Mining Algorithms are Unvalidated

"… formal analysis of the connection between technical statistics and the real world is almost nonexistent."[10]

A recent editorial stated that performance parameters have yet to be established for two methods, the Bayesian confidence propagation neural network (BCPNN) and the multi-item gamma Poisson shrinker (MGPS).[11] This is correct, but newcomers to the field of data mining in pharmacovigilance who may not be familiar with its history should be aware that this is not for lack of trying. For those familiar with the body of work, the following may serve as a useful extension of points previously made. In fact, there is no paucity of validation exercises for all of the commonly used algorithms.[4,12-20] The challenge is not a lack of validation exercises but rather a lack of standards of evidence against which such investigations may be evaluated. The often-cited lack of gold standards for adjudicating causality of spontaneously reported adverse drug reactions is just one obstacle. As shown in table I, there are numerous modifiable parameters and configurations for each data mining algorithm.[4,12-19] The multiplicity of available choices not only complicates the design and interpretation of algorithm validation exercises but even presents considerable opportunities for semantic ambiguity and fundamental miscommunication. The latter confusion is easily avoided by clearly distinguishing between algorithm, metric and threshold when communicating research findings. Regarding the former pitfall, the number of available data mining options and configurations is a double-edged sword consisting of increased exploratory capacity as well as opportunities for false alarms, minimisation of potential true associations and 'self-deception bias'. Self-deception bias can occur when a data miner with a strong incentive to believe in a particular outcome may consciously or subconsciously try to avoid results that contradict pre-existing expectations. By this we mean the data miner may apply nonspecific case definitions of uncertain clinical relevance and/or sequential mining with different subsets of the database and/or candidate data mining parameters, until the 'desired' output is achieved;

Table I. Modifiable parameters and configurations for data mining investigations

Algorithm	Proportional reporting ratios Reporting odds ratios Bayesian confidence propagation neural network (Multi-item) gamma Poisson shrinker Sequential probability ratio tests
Type of database	Public, proprietary (company)
Size of database	35 000–3 000 000 reports Global vs subset of database
Dictionary	MedDRA, COSTART, WHO-ART, local, company developed
Dictionary hierarchy used	PT, HLT, HLGT, SOC
Case definitions	Special search categories (e.g. SMQs) *Ad hoc* case definitions
Drugs	(Suspect) vs (suspect vs concomitant) Certain subsets of drugs removed (to eliminate masking) Grouping by pharmacological/therapeutic class
Study design	Real data vs database simulation Prospective vs retrospective
Methodology	Stratified vs unstratified, reports vs events Binary classifier vs ranking classifier Cross-sectional analysis vs time trend analysis Cumulative vs non-cumulative subsetting of data Deployment in series vs in parallel with other signal detection activities
Performance measures	Sensitivity, specificity, positive predictive values, negative predictive value, ROC curves
Threshold selection/threshold metrics	Disproportionality thresholds/metrics (point vs interval estimates) Case count threshold

COSTART = Coding Symbols and Thesaurus of Adverse Reaction Terms; **HLGT** = higher level group term; **HLT** = higher level term; **MedDRA** = Medical Dictionary for Regulatory Activities; **PT** = preferred term; **ROC** = receiver-operating characteristic; **SMQs** = standardised MedDRA queries; **SOC** = system organ class; **WHO-ART** = WHO Adverse Reaction Terminology.

the equivalent of the classical multiple comparisons. This form of confirmation bias may be especially treacherous for retrospective exercises and *post hoc* interpretations.[21]

It is important to note that published validation exercises typically involve the use of data mining algorithms in isolation. Real-life pharmacovigilance settings involve the use of these algorithms as potential supplements and not as substitutes for traditional signal detection practices based on the clinical and epidemiological domain expertise of the 'prepared mind'.[22] Notably it seems that less attention is directed to standardising and refining core principles of signal detection based on scientific cognition. The complex, subtle and subjective cognitive processes of the 'prepared mind' operating alone or in conjunction with a data mining algorithm defy explicit characterisation. Therefore, although the reported performance of data mining algorithms used in isolation is informative and important, it is not sufficient for fully adjudicating utility in naturalistic pharmacovigilance settings where the increase in predictive value over existing practices obtained by adding algorithms to the pharmacovigilance toolkit is the determinative factor. Other important considerations in the interpretation of validation exercises are that some measures of accuracy, such as sensitivity, specificity and receiver-operating characteristic (ROC) curves that have been reported are very useful but ultimately decouple accuracy from predictive power, which is based on the prevalence of the conditions targeted for screening. It must also be realised that in real life one needs to consider the penalty associated with false positive and false negative findings so that maximising overall accuracy might not be the most desirable goal in all settings.

2. Data Mining Algorithms Allow Data Miners to Objectively Screen Spontaneous Reporting System Data

"… some statistics users (statisticians included) hold the deep conviction that we should let the data

speak for themselves. This conviction is the kernel of naive empiricism that data collection and statistical analysis will automatically lead to scientific discovery."[23]

Another assertion that has appeared in various forms in the published literature[4] and in other professional venues is that some data mining algorithms allow safety reviewers to objectively view or screen large spontaneous reporting system databases. At the present time, there is nothing more objective about the view of spontaneous reporting system databases provided by looking at the database itself versus a set of 'signal scores' derived from that database. The idea that these methods objectively screen large spontaneous reporting system databases is unfounded and predicated on the erroneous idea that subjective judgments are not involved in the construction, application and interpretation of these models and tools. Not only are all probability models subjective but, more importantly, there is currently no theoretical basis or firm empirical support establishing universal threshold cutoffs defining a potential 'signal' (although some have been recommended, e.g. proportional reporting ratio [PRR] ≥2 with χ^2 ≥4, N >2;[2] EB_{05} ≥2, N >0[4]) or the preferred selection(s) from the complex space of available methodological choices available to data miners (table I). Indeed, we are reluctant to use the designation 'signal' for statistical arithmetic devoid of clinical context, preferring the term 'signal of disproportionate reporting' instead. Ultimately, at the desktop, the data miner currently using these techniques to 'screen' spontaneous reporting system databases must make numerous subjective judgements and decisions about which thresholds to apply, which of the numerous candidate data mining configurations to use and what the data mean. Therefore, these tools are promising and can provide useful supplementary numerical information but currently their deployment is inherently subjective in nature.

3. Mathematically More Complex Bayesian Algorithms are Superior to Frequentist Algorithms

"Real problems include elements of uncertainty to the extent that more complex models often fail to improve things."[24]

Another possible pitfall in deploying and/or comparing data mining algorithms is automatically assuming that greater mathematical complexity ensures greater utility and/or that the use of advanced techniques such as Bayesian modelling and shrinkage estimators merely improves the signal-to-noise ratio. Although it is possible that time and experience will bear this out, we caution against 'seduction bias', in which the validity or predictive content may be over-rated because of an elaborate mathematical and graphical framework. Ultimately, spontaneous reporting system data is a convenience sample without a well defined probability structure. As such, the decisions within mathematical frameworks around such data inherently involve a strong element of subjectivity.

A variant of this assertion is that data mining algorithms with a more extensive mathematical framework are preferred, with frequentist algorithms relegated to 'horse and buggy' status. The dialogue about which, if any method, is *the* preferred approach has been vigorous. In hopes of cutting through the confusion we would summarise the comparative performance of these methods with the caveat that Bayesian shrinkage-to-the-null, covariate stratification and/or additional statistical selections may contribute to observed performance gradients in a given situation. When *commonly cited thresholds* are used, the basic performance differentials between frequentist algorithms and Bayesian algorithms used in isolation *as binary classifiers* can be simply stated as follows: (i) in general, frequentist forms of data mining algorithms (e.g. PRRs, reporting odds ratios) seem to highlight a greater number and variety of drug-event associations than Bayesian algorithms (e.g. BCPNN, MGPS); (ii) for drug event-associations that are highlighted by both, frequentist data mining algorithms tend to do so earlier;[25-27] and (iii) many of the additional drug-event associations obtained with frequentist data mining algorithms are due to confounding, reporting artifact or statistical noise and require additional triage criteria for practical implementation (although both are associated with false positive findings). However, many of the 'extra' drug-event associations highlighted by frequentist methods should not be automatically dismissed. For example, because of the coding multiplicities and redundan-

cies characteristic of hypergranular adverse event dictionaries (e.g. the Medical Dictionary for Regulatory Activities [MedDRA]), many of the 'extra' drug-event associations represent adverse events that are clinically equivalent or related to events highlighted by both frequentist and Bayesian forms and may be highlighted only by the former because they represent less common coding variants. We would caution against automatically discounting clinically related events highlighted by only one type of algorithm as representing noise without carefully accounting for the clinical nature of the adverse events under study.

Bayesian algorithms use a mathematical process that factors the overall reporting experience across drugs and events in part to achieve a statistical 'shrinkage' of the highly variable signals of disproportionate reporting associated with low reporting frequency. However, it is unknown to what degree true drug-event associations might be 'shrunk' along with noise, since the mathematical model contains no explicit clinical criteria and homogenises individual reports. Furthermore, from a statistical perspective, a data miner can apply *ad hoc* shrinkage rules to any type of algorithm that will return a score that is adjusted for low reporting frequencies. The performance differentials between shrinkage rules are unknown. Over-attention to statistical differences between variants of disproportionality analysis is probably of limited value since an important performance-limiting *methodological* characteristic of this class of data mining algorithms (ignoring for the moment the enormous reporting artifacts and limitations of spontaneous reporting system data) is that they all operate by projecting high-dimensional data on to 2×2 contingency tables, which can produce misleading associations and mask others that are of interest.

From a broader perspective, expert safety reviewers using non-Bayesian forms of data mining algorithms might integrate clinical judgment and prior knowledge of drugs (drug, therapeutic and pharmacological classes), events, patient populations and diseases to rapidly filter out noise by down-weighting drug-event associations unlikely to be of interest, somewhat like an informal application of Bayesian hypothesis refinement. Such 'clinical shrinkage' allows the safety reviewer using simple forms of data mining algorithms to review a 'wider net' of signals of disproportionate reporting and rapidly discard uninteresting drug-event associations while noting interesting drug-event associations. Therefore, the overabundance of signals associated with simple forms of data mining algorithms may not be prohibitive and the opportunity cost due to lower specificity may not be unreasonable given the increased sensitivity in a given situation. Furthermore, since data mining exercises involve thresholds that are unvalidated, arbitrary and adjustable, performance differentials observed with commonly cited thresholds can be further mitigated or even reversed, when alternative thresholds are selected.[27,28] This underscores our previous statement on the importance of distinguishing between an algorithm and details of its specific implementation in a given data mining exercise. Therefore, no single data mining algorithm has emerged as the method of choice. We caution against drawing universal conclusions because the performance of these methodologies, as mentioned previously, may be highly situation dependent and we view them all as credible options in the pharmacovigilance toolkit, provided their limitations are fully appreciated.

We further note that the aforementioned situation dependence also involves an interplay between data mining procedures and organisational characteristics. Although pharmacovigilance organisations share a common goal, they are not structurally and functionally homogeneous. The importance and impact of the tradeoff between sensitivity and specificity may therefore vary across establishments. As a result, the incremental utility of each algorithm may well depend on the precise mode of deployment as well as the resources, data volume, experience and clinical expertise within the organisation.

4. Data Mining Algorithms are Not Just for Hypothesis Generation

"The familiar caution that correlation does not imply causation has not always been remembered in the enthusiasm inspired by the discovery of apparent relationships."[1]

An insidious correlate of the above phenomena (e.g. seduction bias) that one or more of the authors have observed and has been noted in the published literature[29] is that of 'mission creep', in which

overconfidence in the ability of some algorithms to compensate for the enormous limitations in the data gradually results in claimed applications that are beyond the originally intended use of hypothesis generation. Ambiguous language amplifies the problem. For example, publications report the use of data mining algorithms to '*test* hypotheses', 'identify drug interactions', investigate the '*clinical relevance*' of potential drug interactions and sometimes use terminology such as 'relative risk' to describe disproportionality calculations on spontaneous reporting system data.[8,30] Each of these suggested uses or terminologies for algorithms could be interpreted to imply applications of data mining beyond hypothesis generation. Repeated use of such terms may inadvertently promote premature acceptance of extended applications because they tend to creep, rather than pounce, into the public domain. Although research into extended applications should be encouraged, we wish to emphasise that currently the safest and most appropriate use for these tools is to *help* formulate or refine hypotheses (i.e. is there a drug-event association worthy of further investigation?) and that any unintentional distancing between data and scientific judgement due to the intervening influence of statistical arithmetic carries an associated risk to public safety.

5. Conclusions

"… the quality of the data mining exercise will soon be revealed, in terms of whether the structures which have been unearthed are interesting, valuable, surprising, or previously unknown."[1]

Data mining is a promising adjunct to exploratory data analysis that faces frequent challenges to acceptance. Our position on the level of evidence required to embrace these tools can be summarised by the phrase "trust but verify". Pharmacovigilance is a science and an art and critics must appreciate that much of current pharmacovigilance practice involves non- or semi-quantitative approaches that have not undergone rigorous and systematic validation and could, therefore, be subjected to similar challenges (although there is much more prospective experience with them, unlike data mining algorithms). Particularly, when working with spontaneous data one should not automatically assume that more complex statistical models create significant new understanding of the phenomenon they were designed to illuminate. The incremental utility in naturalistic pharmacovigilance settings is highly contingent on numerous factors. At this point in time, we feel that if a decision is made to use a data mining algorithm, it is best to deploy it in parallel with traditional screening strategies to search for drug-event associations that may have escaped the latter, although the individual characteristics of each pharmacovigilance organisation need to be carefully considered, as discussed previously. In situations in which clinical signal/non-signal is ambiguous, data mining algorithms used in series may be one of many elements that could be used to refine an index of suspicion.

Therefore, we trust that the various forms of disproportionality analysis are credible additions to the pharmacovigilance toolkit and could enhance pharmacovigilance performance in a variety of settings, but they should be verified in direct proportion to the claims that are made about their performance. Our optimism is tempered with a level of caution commensurate with the inherent limitations of the data to which they are being applied and the ensuing potential for misuse of data mining algorithms. We also advise potential users of these algorithms to be mindful of the substantial uncertainties surrounding their use in making qualitative and quantitative inter-drug comparisons.

Each organisation is best qualified to decide what configuration of clinical and computational approaches works best in their respective environment. At the same time, this should not be used as an excuse for not seriously assessing potential gaps in pharmacovigilance practices and considering data mining algorithms as one approach that could improve pharmacovigilance performance. However, the assertion that these tools should be a required element of good pharmacovigilance practice in all settings needs to be empirically verified based on sound rules of evidence. Similarly, if an institution's practice impacts other stakeholders, i.e. health authorities requesting investigations of 'signals', then we must insist on scientific grounds that such actions not be taken on the basis of statistical arithmetic decoupled from clinical judgment and context.

At the time of writing, there is a 'request for proposal' to validate data mining in pharmacovigi-

lance.[31] This will hopefully provide results that will objectively guide discussions about universal policy statements. However, we encourage institutions not to wait for such systematic validations and consider piloting data mining algorithms in their respective institutions, and possibly publish the findings. This will contribute to the collective experience with these tools and we believe that each organisation will soon discover if pharmacovigilance practices are enhanced by their use and, if so, to what degree.

The numerous drug safety professionals, especially those from the US FDA, WHO Collaborating Centre for International Drug Monitoring, Lareb Pharmacovigilance Centre and from drug companies (in addition to the ones we represent) who are referenced in this commentary, are striving to develop, study and test these tools. Although in some instances we may have differences of opinion, their work is of fundamental significance to pharmacovigilance and we extol their achievements in the service of public safety. However, to continue serving public safety and not tangential interests, it is important to neither under-rate nor over-rate any of these methods and the emerging data must be scrutinised. We also caution against overattention to data mining algorithms at the expense of improving data quality and the application of clinical cognition to signal detection and evaluation.

Data mining *"almost spells excitement and opportunity. However, given the difficulties we have outlined above, one should be wary about getting carried away by this. It remains to see precisely who will benefit from data mining activities, beyond companies marketing data mining tools"*.[1]

Acknowledgements

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

References

1. Hand DJ, Blunt G, Kelly M, et al. Data mining for fun and profit. Stat Sci 2000; 15: 111-31
2. Evans SJ, Waller PC, Davis S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. Pharmacoepidemiol Drug Saf 2001; 10 (6): 483-6
3. Van Puijenbroek E, Diemont W, van Grootheest K. Application of quantitative signal detection in the Dutch spontaneous reporting system for adverse drug reactions. Drug Saf 2003; 26 (5): 293-301
4. Szarfman A, Machado SG, O'Neill RT. 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
5. Bate A, Lindquist M, Edwards IR, et al. A Bayesian neural network method for adverse drug reaction signal generation. Eur J Clin Pharmacol 1998; 54 (4): 315-21
6. Bate A, Lindquist M, Orre R, et al. Data-mining analyses of pharmacovigilance signals in relation to relevant comparison drugs. Eur J Clin Pharmacol 2002; 58 (7): 483-90
7. Wilson AM, Thabane L, Holbrook A. Application of data mining techniques in pharmacovigilance. Br J Clin Pharmacol 2004; 57 (2): 127-34
8. 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
9. Hauben M. A brief primer on automated signal detection. Ann Pharmacother 2003; 37 (7-8): 1117-23
10. Wang C. Amoeba regression and time series models. In: Sense and nonsense of statistical inference: controversy misuse and subtlety. New York: Marcel Dekker, 1993: 72-97
11. Lilienfeld DE. A challenge to the data miners. Pharmacoepidemiol Drug Saf 2004; 13 (12): 881-4
12. Hauben M, Zhou X. Quantitative methods in pharmacovigilance: focus on signal detection. Drug Saf 2003; 26 (3): 159-86
13. Kiyoshi K, Daisuke K, Toshiki H. Comparison of data mining methodologies using Japanese spontaneous reports. Pharmacoepidemiol Drug Saf 2004; 13 (6): 387-94
14. Lindquist M, Stahl M, Bate A, et al. 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
15. Van Puijenbroek EP, Bate A, Leufkens HG, et al. A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions. Drug Saf 2002; 11 (1): 3-10
16. Emmanuael R, Tubert-Bitter P, Thiessard F. Evaluation of data mining methods in pharmacovigilance using simulated datasets. Poster presentation at 20th ICPE conference; Bordeaux, France 2004
17. Follmann M, Michel A. Proportional reporting rations for signal detection in the drug safety database of a pharmaceutical company. Poster presentation at 19th ICPE conference; Philadelphia, USA 2003
18. Follmann M, Michel A, Geyer C. Comparison of different methods for signal detection in the drug safety database of a pharmaceutical company. Poster presentation at 20th ICPE conference; Bordeaux, France 2004
19. Yukari K, Eri K, Moriko K. The impact of grouping drugs by ATC codes on detecting a signal from Japanese spontaneous reports. Poster presentation at ICPE Conference; Bordeaux, France 2004,
20. Hauben M, Walsh L, Reich L. Predictive value of a computerized signal detection algorithm (MGPs) when applied to FDA AERS database [abstract]. Pharmacoepidemiol Drug Saf 2005; 14 (S1-S218): S17 (no.35)
21. Hauben M. Application of an empiric Bayesian data mining algorithm to reports of pancreatitis associated with atypical antipsychotics. Pharmacotherapy 2004; 24 (9): 1122-9
22. Trontell A. Expecting the unexpected: drug safety, pharmacovigilance, and the prepared mind. N Engl J Med 2004; 351 (14): 1385-7
23. Wang C. Objectivity, subjectivity, and probability. In: Sense and nonsense of statistical inference: controversy misuse and subtlety. New York: Marcel Dekker, 1993: 137-184
24. Hand DJ. Presentation at joint workshop on statistical data mining; 2003; Eindhoven, The Netherlands, 2003

25. Hauben M, Reich L. Safety related drug-labelling changes: findings from two data mining algorithms. Drug Saf 2004; 27 (10): 735-44
26. Hauben M, Reich L. Drug-induced pancreatitis: lessons in data mining. Br J Clin Pharmacol 2004; 58 (5): 560-2
27. Moseley J, Heeley E, Ekins-Daukes S, et al. Preliminary comparison of 2 signal detection methodologies in the UK regulatory authority spontaneous ADR database. Drug Saf 2004; 27 (12): 950-1
28. Hauben M, Reich L, Gerrits C. Comparative performance of proportional reporting ratios (PRR) and multi-item Gamma Poisson shrinker (MGPS) for the identification of crystalluria and urinary tract calculi caused by drugs [abstract]. Pharmacoepidemiol Drug Saf 2005; 14 (S1-S218): S7
29. Strom BL. Evaluation of suspected adverse drug reactions. JAMA 2005; 293 (11): 1324-5
30. Almenoff JS, DuMouchel W, Kindman LA, et al. Disproportionality analysis using empirical Bayes data mining: a tool for the evaluation of drug interactions in the post-marketing setting. Pharmacoepidemiol Drug Saf 2003; 12 (6): 517-21
31. PhRMA. Request for proposal: validity and value of data mining methods as an adjunct to traditional methods for detecting safety signals from spontaneous reporting databases [online]. Available from URL: http://www.phrma.org/publications/publications/Data_Mining_RFP.pdf [Accessed 2005 Feb 5]

Correspondence and offprints: Dr *Lester Reich*, Risk Management Strategy, Pfizer Inc, 235 E. 42nd Street, New York, NY 10017, USA.
E-mail: Lester.Reich@Pfizer.com