SHORT COMMUNICATION

Drug Safety 2002; 25 (6): 415-421
0114-5916/02/0006-0415/$25.00/0



# Statistical Techniques for Signal Generation
## The Australian Experience

*Patrick Purcell* [1] and *Simon Barty* [2]

1 Adverse Drug Reactions Unit, Therapeutic Goods Administration, Woden, Australian Capital Territory, Australia
2 School of Information Technology and Mathematical Sciences, University of Ballarat, Mt Helen, Victoria, Australia

## Abstract

National voluntary reporting systems generate large volumes of clinical data pertinent to drug safety. Currently descriptive statistical techniques are used to assist in the detection of drug safety 'signals'. Australian data have been coded according to guidelines formulated almost 30 years ago and which have resulted in many drugs which are not associated with an adverse drug reaction or 'innocent bystander' drugs being recorded as 'suspected' in individual reports. In this paper we explore the application of an iterative probability filtering algorithm titled 'PROFILE'. This serves to identify the 'signals' and remove the 'innocent bystander' drugs, thus providing a clearer view of the drugs most likely to have caused the reactions. Reaction terms analysed include neutropenia, agranulocytosis, hypotension, hypertension, myocardial infarction, neuroleptic malignant syndrome, and rectal haemorrhage. In this version of PROFILE, Fishers exact test has been used as the statistical tool but other methods could be used in future. Advantages and limitations of the method and its assumptions are discussed together with the rationale underlying the method and suggestions for further enhancements.

It is possible to construct $2 \times 2$ contingency tables for a drug/reaction association. Our initial attempts at signal generation by statistical assessment of Australian voluntary reporting data using rate ratios and odds ratios from these contingency tables met with only partial success. Regardless of whether rate ratios or odds ratios were used, the problem of defining an objective threshold remained. Any given threshold resulted in unacceptable false positive and false negative results when compared with the product information and published literature. Zero values in the denominator often precluded use of odds ratios. Moreover the data are influenced by confounding and biases.

Sources of variability in the Australian data are likely to include the background reporting rate, reaction type (including the system organ class involved, incidence and seriousness), drug status and usage and variations in these with the passage of time. Factors influencing the decision to report an individual case also include awareness of the possibility of adverse drug reactions (ADRs), the level of communication between patient and doctor, and publicity. However, the intrinsic nature of the re-

action itself and its impact on the patient seem likely to have a major influence. For example it is much easier to recognise a rash as an ADR than, say, raised intracranial pressure, aseptic meningitis or pancreatitis. This suggests that the data should be analysed by individual reaction terms to accommodate differences in the recognition and reporting of various reaction terms.

For any given reaction term it is common to find that a relatively small number of drugs account for the majority of drug/reaction associations in the raw data. For example drugs with three or more reports of the association may account for the major portion of the data but only a minority of the drugs listed against the reaction in the raw data.

Two further factors need to be considered. Firstly, many drug safety signals involve a relatively small number of reports associating a drug with the reaction. Often this may be five or fewer reports. Secondly, the causality guidelines that have been used for more than 30 years for Australia's voluntary reporting scheme often result in several drugs being simultaneously suspected in individual reports. These guidelines require the recording as 'suspected' of drugs other than those designated by the reporter where the drugs cannot be excluded on the basis of the information in the report. Thus, for many individual reports several drugs will be recorded as suspected on the basis of a common temporal association. For example digoxin, furosemide (frusemide) and potassium supplements are commonly 'implicated' in this context. This deliberate policy was intended to enhance sensitivity, reduce bias and allow for possible interactions. The effects of this policy on the data have never been formally evaluated. Hence, in many individual reports it is difficult to determine which of several suspected drugs is the likely cause of the adverse reaction. To date no systematic solution has been available to identify these 'innocent bystanders'.

| | |
|---|---|
| A | B |
| C | D |

**Fig. 1.** A 2 × 2 contingency table. **A** = reports connecting the drug of interest with the specified reaction; **B** = reports of the drug of interest with all other reactions in the relevant 'system organ class' (SOC); **C** = reports of the specified reaction with other drugs received over the life of the drug; **D** = reports of all other drugs with all other reactions in the relevant SOC over the life of the drug.

## 1. Setting, Materials and Methods

The Australian database of suspected ADRs is used as the source of voluntary reporting data received from health professionals in both hospital and community settings and pharmaceutical companies. Raw data were extracted for each specific reaction term and the analysis was stratified to the relevant WHO 'system organ class' (SOC).

Accordingly the 2 × 2 contingency table recently adopted for use by the UK Medicines Control Agency[1] was modified for analysis of these data as shown in figure 1.

It was thought that variation due to external sources might be reduced if parameters were calculated for a given SOC (rather than all reactions) and the data limited to the period of reporting for the drug in question.

## 2. The PROFILE Analysis

All reports for a particular reaction term were analysed simultaneously. The familiar 2 × 2 contingency table was stratified to the SOC relevant to the reaction being analysed and for each drug associated with the reaction the data were limited to the period of monitoring of the drug. Fisher's exact one-tailed right was calculated for each drug associated with the reaction. In the context of the Australian voluntary reporting data, Fisher's exact p correlated with the standardised normal deviate

(for A >2 as defined in figure 1) and also correlated with the corresponding Poisson probability. The drug with the lowest Fisher's exact p was identified. This drug was then inferred to be the single most likely cause for all the reports in which it is associated with the reaction of interest. Other drugs also suspected in the same group of reports implicating that drug are then regarded as 'innocent bystanders' and their tallies were reduced accordingly (the A cell was reduced and the C cell increased by the same amount; see figure 1). All the remaining reports of the reaction (implicating other drugs independently of the first drug) were then analysed similarly until the iterative process was completed and the number of 'drug suspects' exactly matched the number of reports of the reaction. The process is repeated until each individual report has been attributed to a single drug.

In essence the raw data were used to construct a 'probability ladder' which was then used to progressively filter out the 'innocent bystander' drugs. This method has been named the PROFILE analysis. The acronym stands for (top-down) PRObability FILtEring. The algorithm has been programmed in SAS software. Drugs listed in the raw data are then classified as either PROFILE positive (surviving the analysis) or PROFILE negative (regarded as not being causally associated).

## 3. The Peto Odds Ratio

The PROFILE method has been tested in a three way comparison with: (i) drugs ranked according to whether they were 'first named' in the reports. It was considered that, of the suspected drugs, the first named would usually be the one suspected by the reporter. So the first named drug was used as a surrogate for the drug designated by the reporter; (ii) drugs ranked according to the PROFILE method using Fisher's exact p; and (iii) drugs ranked according to a modified PROFILE method in which the Peto odds ratio was used as an alternative to Fisher's exact p. The Peto odds ratio is the ratio of the difference between observed and expected values to the hypergeometric variance or

$$\text{Peto Odds Ratio} = \frac{\text{Observed} - \text{Expected}}{N(N^2 - 1)}$$

where the expected value is

$$\frac{(A+B) \times (A+C)}{N} \text{ and } N = A + B + C + D$$

Whereas the first two methods produced results which were in substantial agreement, the last method produced very idiosyncratic results and so was not pursued further.

## 4. Comparison of z Statistic and Fisher's Exact Test

When Fisher's exact p was plotted against the corresponding p(z) for PROFILE positive data, the correlation was acceptable only if a threshold of A > 2 (as defined above) was applied. Sigmoid curves (reflecting a 'probability spectrum' as shown in figure 2) were obtained when the PROFILE positive data were used to plot Fisher's exact p

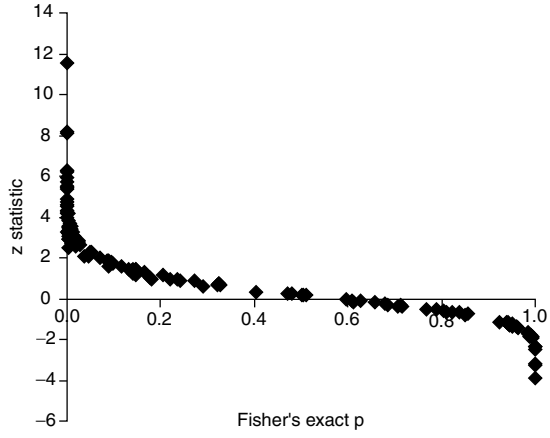


**Fig. 2.** Each datapoint represents a drug that has survived PROFILE analysis with three or more reports of hypertension. Strong, intermediate and weak statistical associations are evident reflecting a 'probability spectrum' as described in the text. This sigmoid shape is typical of most reactions analysed by the PROFILE method.

against the standardised normal deviate (z statistic). With minor variations, the sigmoid pattern was reproducible for the various reactions. It was noted that the curves were cleaner if drugs with fewer than three reports were excluded. These may be either 'noise' or 'evolving signal' but the distinction cannot be made statistically. In practice it is generally considered unacceptable to rely on associations with fewer than three reports.

## 5. Results

Outputs were examined for the following reaction terms:

- Neutropenia: Clozapine (z = 7) was ranked first in the PROFILE positive list.
- Agranulocytosis: Clozapine, despite being an acknowledged cause of agranulocytosis with 60 reports, was lowly ranked (z = –1.17 and Fisher's exact p = 0.94) This is probably due to the white cell monitoring and intervention which is an integral part of clozapine use. This has generated large numbers of clozapine reports in system organ class 1220 thereby weakening the statistical association. Thus, neutropenia is detected and the drug is stopped in many cases preventing agranulocytosis. Carbamazepine (z = –3.8, Fisher's exact = 1.0) was also lowly ranked although it is acknowledged to cause agranulocytosis 'very rarely'. Aspirin (acetylsalicylic acid) [17 reports], digoxin (10), ampicillin (8) and diazepam (7), were identified as 'innocent bystanders' (PROFILE negative).
- Hypotension: Polygeline, albumin, streptokinase, and thiopentone were ranked highly in the PROFILE positive list. However, sildenafil, felodipine and amlodipine were lowly ranked ($z < -3.3$ and $p = 1.0$) reinforcing the conclusion that there is no threshold in the probability spectrum in this context.
- Hypertension: Phenylpropanolamine, tranylcypromine, clonidine, phenelzine and ergometrine were highly ranked ($z > 7.8$). Clozapine, ginkgo biloba and moclobemide all had $z > 4.6$. The serotonin reuptake inhibitors (SSRIs) fluoxetine, paroxetine and sertraline were all lowly ranked ($z < -0.5$ and $p > 0.75$). It is important to note that drugs cannot not be discounted because of their lowly ranked position in the 'probability spectrum'.
- Myocardial infarction: The PROFILE positive list revealed significant confounding by indication. Drugs used for treatment of hypertension, hyperlipidaemia, AIDS, type 2 diabetes mellitus, angina, smoking cessation, and thrombotic disorders featured prominently. Sildenafil headed the list – sexual intercourse being associated with myocardial infarction – 'success' being a double edged sword in this instance!
- Neuroleptic malignant syndrome (NMS): Haloperidol, fluphenazine, clozapine, risperidone, pericyazine, thioridazine, olanzapine, levodopa-carbidopa, chlorpromazine, droperidol, trifluoperazine and metoclopramide were all strongly associated with NMS. The PROFILE negative list is of interest because it included benztropine (13 reports) which is used for treating hypertonia (and therefore likely to be 'noise'), and lithium carbonate (10 reports) which is known to interact with antipsychotic agents in producing NMS. Hence the PROFILE negative list may contain drugs which interact with one or more drugs on the PROFILE positive list.
- Rectal haemorrhage: The PROFILE positive list revealed groups of drugs and possible mechanisms – warfarin (an anticoagulant), antibiotics (pseudomembranous colitis), nonsteroidal anti-inflammatory drugs which are known for their ulcerogenic effects on gastrointestinal mucosae, drugs associated with thrombocytopenia, other drugs associated with platelet dysfunction (SSRIs, aspirin).

In summary, the effect of the PROFILE analysis using Fisher's exact p was averaged across 6629 reports which included mention of seven chosen reaction terms. The raw data comprised almost 2000 drugs linked with 8000 associations. The PROFILE analysis identified about 17% of the as-

sociations as 'noise' due to co-suspected drugs. If the 'signal' is defined to be three or more reports surviving PROFILE then over 80% of the reports can be accounted for by about 25% of the drugs originally suspected in the raw data. That is, 75% of those drugs are either 'noise' or 'evolving signal'.

## 6. Detection of Quality Control Problems with Medicines Using Fisher's Exact p

Two examples of quality control problems with medicines were briefly considered. In each case it was demonstrated that the application of Fisher's exact p would have generated an early statistical signal. Hypotension specifically related to a problem with some batches of polygeline would have been identified as a signal on receipt of the third such report over a 3-month period. The receipt of the first six reports of back pain with a particular batch of intravenously administered fluorescein within a week would have generated an immediate statistical signal (in this context defined as a consistent decline in Fisher's exact p over time).

## 7. Discussion

Fisher's exact p is especially appropriate when dealing with small numbers. However, its use for ranking purposes in this context can be criticised because the assumption of independent tables has been violated. Notwithstanding this, the method appears to function satisfactorily with Australian data. Moreover, it can be demonstrated that for the Australian data there is a close correlation between Fisher's exact probability and the corresponding Poisson probability. Future research should examine the use of Poisson probability as the statistical tool for use with PROFILE.

The PROFILE method as presented here is not reliant on assumptions about the frequency distribution of the data, requires no calculations of confidence intervals, tests of statistical significance, probability thresholds, or utilisation data. The only major assumption is that for each individual report there is a single most likely cause (which is least likely to be explained by chance) and that other drugs suspected in the same reports are in fact 'innocent bystanders' unlikely to be causally implicated in the reaction. This method is not appropriate for detection and analysis of interactions.

By removing a major component of 'noise' the PROFILE method provides a clearer view of the data. It can provide useful insights but is not a panacea. It has revealed both positive and negative 'biases' in the data. The concept of a probability threshold has proven to be untenable in this context. Empirically, it seems likely that, for the time being at least, a minimum of three reports surviving PROFILE analysis will be the working definition of a 'statistical signal' for Australian data. The PROFILE POSITIVE lists will include some confounding by indication and are also likely to retain some residual 'noise' due to chance co-morbidity. It should be remembered that some drugs will survive the PROFILE analysis by default in cases where no other drugs have been suspected, either because the reporter was not aware of them or did not consider them. This is particularly important in the circumstance where a drug survives the analysis with only one or two reports. Nonetheless, the reduction of confounding by other, simultaneously suspected drugs should greatly facilitate recognition of potential new signals. Use of the method should permit valid comparison of Australian data with those of other nations analysed with the Bayesian probability method devised by Edwards et al.[2-4]

The PROFILE negative lists are likely to contain some drugs which have interacted with drugs in the PROFILE positive lists. In such cases the primary drug (surviving the analysis) may be viewed as a 'necessary cause' and the other interacting drug as a possible 'contributory cause' as described by Meyboom et al.[5,6]

The final list of drugs produced by this method is more informative than the original computer generated printout of raw data. This information

derived from probability filtering complements the qualitative information available through traditional methods of assessment of adverse drug reaction reports. The two methods are not mutually exclusive and statistical signals should not be interpreted in isolation.

Much more work will be needed to validate the method. It should always be remembered that statistical significance does not necessarily equate with clinical significance. It should also be noted that the quality of reports is critical to their interpretation. An important caveat is that a statistical association can be 'diluted' or weakened by high volume reporting (for the specific drug) within the relevant SOC.

The PROFILE method may be particularly useful in Australia in that it permits objective inter-drugs comparison which may not otherwise be valid due to the nature of Australian voluntary reporting data. A major finding of this analysis is that the raw Adverse Drug Reactions Advisory Committee (ADRAC) data include many 'innocent bystander' drugs attributable to Australia's coding practice which permits multiple drugs to be suspected in the same report. In light of the findings of this analysis, Australian coding policy is being reviewed. Also, for any given reaction term, the drugs associated with that reaction can be viewed statistically as part of a continuous 'probability spectrum' and it is possible to have genuine signals with weak statistical associations. Fisher's exact probability could also be used to monitor short-term increases in frequency of reports of reactions associated with quality control problems requiring regulatory action.

## 8. Challenges for the Future

The ADRAC software is currently being upgraded. Hopefully this will soon permit electronic submission of reports, greater access to and better use of the data. Importantly it should also facilitate earlier signal recognition and investigation. The impact of the use of the Medical Dictionary for Regulatory Activities (MedDRA), which is the terminology now used by the Therapeutic Goods Administration/ADRAC, is difficult to predict.

There is a need for automated techniques which will assist in detecting quality control problems with medicines. New and innovative techniques will also be needed for the detection of interactions. Examining the evolution of signals over time may provide us with insights into the impact of publication of signals. Identification and measurement of biases should help us to interpret the data. Identification of risk factors and confounding by indication also need to be addressed systematically. In addition to stratifying the data by system organ class, simultaneous stratification by drug type using the anatomical therapeutic chemical (ATC) classification codes needs to be considered as this may produce more specific and clinically relevant comparisons thereby minimising the effect of confounding by indication. Undoubtedly some of these will be investigated by the University of Ballarat, Victoria, Australia.

## Acknowledgements

The authors wish to thank the Therapeutic Goods Administration (TGA) and the Adverse Drug Reactions Advisory Committee (ADRAC) for granting permission to use the Australian voluntary reporting database and related software. For their helpful comments on earlier versions of the manuscript, we would like to thank Professor John McNeil, Department of Epidemiology and Preventive Medicine, Monash University, Professor Peter Pillans, Department of Clinical Pharmacology and Toxicology at the Queen Elizabeth Hospital, Brisbane, Professor Ric Day, Department of Clinical Pharmacology at St Vincent's Hospital, Sydney, Professor John Marley, Pro-Vice Chancellor of the Faculty of Health, University of Newcastle (Australia), and Dr John McEwen, Head of the Adverse Drug Reactions Unit of the TGA.

The views expressed are those of the authors and do not necessarily represent those of the TGA or ADRAC.

## References

1. Evans SJW, Waller P, Davis S. Proportional reporting ratios: the uses of epidemiological methods for signal generation. Pharmacoepidemiol Drug Saf 1998; 7 Suppl. 2: S102

2. 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: 315-21
3. Bate A, Lindquist M, Orre R, et al. Identifying and quantifying signals automatically [abstract]. Pharmacoepidemiol Drug Saf 1998; 7 Suppl. 2: S103
4. Lindquist M, Edwards IR, Bate A, et al. From association to alert: a revised approach to international signal analysis. Pharmacoepidemiol Drug Saf 1998; 7: S15-25
5. Meyboom RHB. Signal assessment: detecting adverse drug reactions: pharmacovigilance in the Netherlands. The Netherlands Pharmacovigilance Foundation, LAREB, 1998 Oct
6. Meyboom RHB, Egberts ACG, Edwards IR, et al. Principles of signal detection in pharmacovigilance. Drug Saf 1997 Jun; 16 (6): 355-65

Correspondence and offprints: Dr *Patrick Purcell*, Adverse Drug Reactions Unit, Therapeutic Goods Administration, P.O. Box 100, Woden, ACT 2606, Australia.
E-mail: patrick.purcell@health.gov.au