PHARMACOEPIDEMIOLOGY AND DRUG SAFETY 2002; **11**: 3–10
Published online 6 February 2002 in Wiley InterScience (www.interscience.wiley.com). **DOI**: 10.1002/pds.668

ORIGINAL REPORT

A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions

Eugène P. van Puijenbroek*[1], Andrew Bate[2,3], Hubert G. M. Leufkens[4], Marie Lindquist[2], Roland Orre[5] and Antoine C. G. Egberts[4,6]

[1]*Netherlands Pharmacovigilance Foundation Lareb, 's-Hertogenbosch, the Netherlands*
[2]*The Uppsala Monitoring Centre, Uppsala, Sweden*
[3]*Department of Clinical Pharmacology, Umeå University, Sweden*
[4]*Department of Pharmacoepidemiology and Pharmacotherapy, Utrecht Institute for Pharmaceutical Sciences, Utrecht, the Netherlands*
[5]*Department of Mathematics, Stockholm University, Sweden*
[6]*Hospital Pharmacy Midden-Brabant, TweeSteden Hospital and St Elisabeth Hospital, Tilburg, the Netherlands*

SUMMARY

Purpose A continuous systematic review of all combinations of drugs and suspected adverse reactions (ADRs) reported to a spontaneous reporting system, is necessary to optimize signal detection. To focus attention of human reviewers, quantitative procedures can be used to sift data in different ways. In various centres, different measures are used to quantify the extent to which an ADR is reported disproportionally to a certain drug compared to the generality of the database. The objective of this study is to examine the level of concordance of the various estimates to the measure used by the WHO Collaborating Centre for International ADR monitoring, the information component (IC), when applied to the dataset of the Netherlands Pharmacovigilance Foundation Lareb.

Methods The Reporting Odds Ratio − 1.96 standard errors (SE), proportional reporting ratio − 1.96 SE, Yule's Q − 1.96 SE, the Poisson probability and Chi-square test of all 17 330 combinations were compared with the IC minus 2 standard deviations. Additionally, the concordance of the various tests, in respect to the number of reports per combination, was examined.

Results In general, sensitivity was high in respect to the reference measure when a combination of point- and precision estimate was used. The concordance increased dramatically when the number of reports per combination increased.

Conclusion This study shows that the different measures used are broadly comparable when four or more cases per combination have been collected. Copyright © 2002 John Wiley & Sons, Ltd.

KEY WORDS — pharmacovigilance; disproportionality; quantitative signal detection

INTRODUCTION

When medicinal products are marketed, case reports of suspected adverse drug reactions (ADRs) are reported to spontaneous reporting systems on a national level. One task is to detect and investigate possible new side-effects of these drugs. All case reports are filed in databases at the National Centres as well as sent onto the WHO Collaborating Centre for International Drug Monitoring (the Uppsala Monitoring Centre).[1] Usually, trained assessors regularly examine every incoming reported combination between a drug and a suspected ADR for possible signals in a case by case analysis. A systematic

* Correspondence to: Dr E. P. van Puijenbroek MD, PhD. Netherlands Pharmacovigilance Foundation Lareb, Goudsbloemvallei 7, 's-Hertogenbosch the Netherlands. Tel: (+)31 73 6469707. Fax: (+)31 73 6426136. E-mail: e.vanPuijenbroek@Lareb.nl

Received 4 December 2000
Revised 26 March 2001
Accepted 12 September 2001

continuous review of the combinations present in the database is necessary to optimize the primary goal of spontaneous reporting systems, i.e. monitoring for unexpected or unknown ADRs or signal detection.[2] The WHO defines a signal as: 'Reported information on a possible causal relationship between an adverse event and a drug, of which the relationship is unknown or incompletely documented previously'. Often, a limited number of reports represent a signal.[3] Because of increasingly large numbers of case reports being stored in databases, adequate signal detection without automated quantitative screening is becoming time consuming and inefficient, because of the sheer load of information to be assessed. In quantitative signal detection, combinations of a drug and a clinical event that are disproportionately highly represented in the database, may represent an important signal based upon a difference from the background frequency.[4] Subsequently these combinations must still be analysed and interpreted by the critical human mind. In contrast to hypothesis testing where quantitative estimates are used to express the frequency of a signal, in spontaneous reporting systems they are used to determine the probability of a combination being a signal or not, based on disproportionate reporting.

The use of a measure of disproportionality is currently applied in various national spontaneous reporting centres as well as in the Uppsala Monitoring Centre. Several point estimates like the Reporting Odds Ratio (ROR), Proportional ADR Reporting Ratio (PRR) or Yule's Q, have been used, in combination with additional estimators of the precision of point estimates[5–9] such as the Chi-square test, or the lower limits of the 95% confidence intervals of the point estimates. Furthermore, the chance of the number of reports being reported on a certain combination, under the assumption that no relationship exists between the reported suspected ADR and the suspected medication, can be calculated by means of the Poisson probability.[10,11] Another approach is the use of Bayesian logic, specifying the relation between the prior and posterior probability before and after linking data fields, and of adding new data to the database, currently being used for example by the Uppsala Monitoring Centre in the Bayesian Confidence Propagation Neural Network analysis (BCPNN).[12–14] This relationship is expressed as the 'information component'(IC). Also the FDA is developing a Bayesian approach.[15]

The aim of this study is to examine the concordance of the various estimates, and to clarify the way they are related to each other, when applied to the dataset of suspected ADRs reported to the Netherlands Pharmacovigilance Foundation Lareb. Calculations of measures of disproportionality are primarily based upon a two-by-two contingency table (Figure 1). Since all the measures of disproportionality are based on the same principles of calculation using the 2×2 table, results should be closely concordant. On the other hand it is important to know how the different methods perform in practice, particularly at low numbers of reports of a particular combination.

There is no true 'gold' standard to compare methods, but the BCPNN has been tested for performance for signal determination against standard literature sources on a retrospective basis.[14] In order to obtain maximum information from the comparison, it was decided to express the level of concordance of the other methods with the IC − 2 SD in terms of sensitivity, specificity, positive predictive value and negative predictive values, instead of using a measure of concordance such as the kappa statistic. It is appreciated that the use of 'sensitivity' and 'specificity' under these circumstances can be misleading, and these terms should be seen in a relative sense.

METHODS

The dataset of the Netherlands Pharmacovigilance Foundation Lareb was used for the analysis. Lareb maintains the spontaneous adverse drug reaction reporting system in the Netherlands on behalf of the Dutch Medicines Evaluation Board.[16] All reports received by 1 January 2000 were included in the analysis.

Based on the 2×2 table and with respect to the background frequency of associations of drugs and suspected ADRs in the database, the following point and precision estimates were calculated for all combinations: IC − 2 SD, ROR minus 1.96 standard error (SE), PRR − 1.96 SE, Yule's Q − 1.96 SE, Chi square (with Yates correction) and the Poisson probability. In

	Reports with the suspected ADR	Reports without the suspected ADR
Reports with the suspected drug	a	b
All other reports	c	d

Figure 1. Two-by-two contingency table (for corresponding formulas see Appendix A)

the event the measures could not be calculated for mathematical reasons, the missing combinations were excluded from the analysis. For the corresponding formulas see Appendix A. The concordance of the different measures of disproportionality were compared with the results of the BCPNN analysis as reference measure. The BCPNN analysis calculates IC and IC − 2 SD values for all drug–ADR combinations in its routine use. Those combinations that have IC minus 2 standard deviation (SD) greater than zero are highlighted for review.[12–14,17] In spontaneous reporting, a fairly small number of reports may be sufficient to generate a signal.[18,19] Furthermore, for the use of the various approaches described, a certain statistical distribution is assumed. Since it is not clear whether these assumptions are always fulfilled, calculations might be less appropriate in the event of small numbers. For both reasons, the concordance of the various tests with the results of the BCPNN as the number of reports per combination varied, was also studied. In this respect, separate calculations were made for those situations in which number of reports per combination was greater or equal to two, three, four or six. Only combinations with a minimum amount of two reports were selected, since the use of a measure of disproportionality was not considered useful in the event that only one report has been received, even though measures of disproportionality can be calculated in the latter situation.

For ROR − 1.96 SE, PRR − 1.96 SE, Yule's Q − 1.96 SE, Chi square (with Yates' correction) and the use of the Poisson probability, sensitivity, specificity, positive predictive value and negative predictive value as well as the percentage of cases in which the various measures could be calculated, were determined with IC − 2 SD as the reference measure. A positive association was defined as IC − 2 SD > 0 and negative as IC − 2 SD < 0. For calculating point and precision estimates, including the IC and IC − 2 SD and the indicators of the concordance, Microsoft Excel 97 was used.

RESULTS

On 1 January 2000, 26 555 reports were filed in the database of the Netherlands Pharmacovigilance Foundation Lareb. These reports involved a total number of 39 790 reported suspected adverse drug reactions (ADRs) which concerned 17 330 different combinations between a drug and a suspected ADR. The number of combinations with one, two, three and four or more reports was 11.856, 2.455, 1.072 and 1.947 respectively. The mean number of reports per combination was 2.3.

Table 1. Sensitivity, specificity, positive predictive value and negative predictive value concerning the use of various point estimates and tests in comparison with IC − 2SD, regarding different numbers of reports per combination. Also the percentage of combinations for which a point estimate could be calculated, is provided

Test	Number of reports	Sensitivity	Specificity	Positive predictive value	Negative predictive value	Percentage calculated
ROR − 1.96 SE > 1	a ≥ 2	1.00	0.66	0.41	1.00	94.8
	a ≥ 3	1.00	0.77	0.70	1.00	99.9
	a ≥ 4	1.00	0.84	0.83	1.00	100.0
	a ≥ 6	1.00	0.89	0.91	1.00	100.0
PRR − 1.96 SE > 1	a ≥ 2	1.00	0.70	0.44	1.00	94.8
	a ≥ 3	1.00	0.81	0.73	1.00	99.9
	a ≥ 4	1.00	0.86	0.85	1.00	99.9
	a ≥ 6	1.00	0.88	0.90	1.00	100.0
Yule's Q − 1.96 SE > 1	a ≥ 2	1.00	0.60	0.38	1.00	99.5
	a ≥ 3	1.00	0.73	0.67	1.00	99.9
	a ≥ 4	1.00	0.81	0.80	1.00	99.9
	a ≥ 6	1.00	0.86	0.89	1.00	100.0
Chi square	a ≥ 2	1.00	0.71	0.46	1.00	100.0
(Yates correction)	a ≥ 3	1.00	0.80	0.73	1.00	100.0
($p < 0.05$)	a ≥ 4	1.00	0.83	0.82	1.00	100.0
	a ≥ 6	1.00	0.85	0.88	1.00	100.0
Poisson	a ≥ 2	1.00	0.53	0.34	1.00	100.0
($p < 0.05$)	a ≥ 3	1.00	0.66	0.61	1.00	100.0
	a ≥ 4	1.00	0.74	0.75	1.00	100.0
	a ≥ 6	1.00	0.79	0.84	1.00	100.0

The results of the comparisons for different numbers of reports per combination are presented in Table 1. In the majority of cases a combination of point and precision estimate could be calculated. In the event two or three cases per combination were reported, calculating the ROR, PRR and ROR − 1.96 SE was not possible in about 1–5% of the cases. In general sensitivity was high but the specificity was rather low for all measures applied. In all situations the concordance of combination of point estimate and precision estimate increased dramatically in case the number of reports per combination increases. In the event that four or more reports were received on a combination, all approaches gave comparable results.

DISCUSSION

In comparison with IC − 2 SD, the various methods were highly sensitive but had a rather low specificity. Only when the Poisson probability and Chi square were used, could all combinations be analysed. Nevertheless, in the dataset used in this study, for instance the ROR − 1.96 SE could be calculated for 99.95% of all the drug–ADR combinations reported three or more times.

When four or more reports per combination were present, no clear differences were found between the use of the various measures, and the use of the IC − 2 SD. Although the percentage of combinations in the Lareb dataset with four or more reports received is rather low (11.2%), this subset is of particular interest for signal detection.

For drug–ADR combinations in the dataset which are listed on less than four reports, results of the various analyses show some differences. For all measures sensitivity is still high with respect to IC − 2 SD, but specificity rapidly declines for such a low number of reports. This implies that all combinations highlighted as potential signals by use of IC − 2 SD are also highlighted by the other measures under investigation, whereas not all combinations highlighted by the other measures have a positive IC − 2 SD. So, either the number of false positive signals increases for combinations of less than four reports, for each of the measures compared to the IC − 2 SD, or the potential positives highlighted by the other measures are in fact true positives which the BCPNN might go on to highlight later, as more information accumulates. Further detailed investigation is needed to attempt to determine which of these scenarios is most likely. This evaluation is made harder due to the lack of a true gold standard for discrimination of true and false signals. Thus careful ongoing evaluation of how these potential signals develop over time may be the most appropriate method of investigation. When the number of combinations for which a measure could be calculated was not 100%, the missing combinations were excluded from the analysis. These combinations should also be checked by other techniques to determine if they might represent true signals.

For all tests, results are more comparable with IC − 2 SD when the number of reports per combination increases. Although not presented here, other measures such as the Poisson probability were also studied as the reference measure, but whichever method was used for comparison, poor concordance was found at low counter values. This lack of concordance may be explained by the fact that for a small number of reports the assumed type of distribution of the various classical methods (e.g. Gaussian or Poisson distribution) will have a strong influence. For the tests used, some basic assumptions should be applied. For instance, concerning the use of Chi square, on tables with more than a single degree of freedom, a minimum expected frequency of 5 can be regarded as adequate. If there is only one degree of freedom (which is the case in our two-by-two contingency table), a minimum expected frequency of 10 is much safer.[20] When the expected numbers are small, but greater than 5, another option is to apply continuity correction (Yates' correction).[21] Even so the results of the test should be interpreted cautiously. In general only a small number of reports per combination is necessary to trigger a signal. In this situation, calculation of the Poisson probability therefore is safer, although when using a Bayesian implementation, the IC can also be calculated for small numbers. Calculating the confidence interval of the odds ratio and Yule's Q is also subject to limitations. For small numbers of reports the distribution may be skewed, and calculations based upon a Gaussian distribution cannot be applied without caution.

The prior assumption for the Bayesian derivation of the IC is of independence between the drug and ADR. This causes a dampening effect of the IC at very low numbers of cases, making positive ICs less positive and negative ICs less negative than might be envisaged looking purely at the proportion of expected/observed cases.

These differences between the quantitative measures serve to reiterate the crucial importance of clinical and pharmaceutical information, as well as other data in signal detection.[3] In the event of small numbers of cases other aspects contribute increasingly in the selection of signals, such as the clinical

information available and the level of documentation of the reports as the potential signal can be analysed on a case-by-case basis. Any combination selected in spontaneous reporting databases using purely statistical methods, should be carefully evaluated and confirmed by other means before making any kind of regulatory decision.[22] Additional analysis of the signals, i.e. by evaluation of the original reports, is therefore warranted.[23–25]

Methodological considerations

Although in quantitative signal detection no true gold standard is available, we have chosen the information component of the WHO as being the reference measure for the following reasons. Firstly, when used for analysing the WHO database, this approach yielded a positive predictive value of 44% and a negative predictive value of 85% in the detection of signals as compared with reference literature sources.[13] In this study the authors discuss the difficulty of defining a gold standard in signal detection. We considered that the availability of the performance information in that study, albeit subject to limitations, would allow the results of the current study to be placed in a useful context. Secondly, in contrast to other measures, both point estimate (IC) and its probability interval can be calculated under all circumstances. The confidence intervals for the ROR, PRR, Yule's Q were calculated as the standard error, but in the Bayesian approach, the standard deviation was calculated from the IC distribution. Furthermore, for calculating the lower limit of the confidence interval IC − 2 SD instead of IC − 1.96 SD has been used since it is routinely implemented in this way, and was used for the retrospective evaluation study outlined above.

An alternative method of comparison could have been used such as the kappa statistic, but a drawback is that this does not distinguish between a situation of high sensitivity and low specificity, and one of poor sensitivity and high specificity.

In the Poisson probability or the Chi-square test only the chance that the observed frequency differs from the expected frequency is provided. This situation differs from the other tests like Yule's Q − 1.96 SE, ROR − 1.96 SE and PRR − 1.96 SE where the prior assumption is made that combinations we are looking for, occur more often than the expected frequency. If we wanted to look for combinations that occur less frequently in combination with a certain drug we should use for instance ROR + 1.96 SE. In our analysis we did not take these differences into account.

Choice of a measure of disproportionality

In this study the level of concordance between the different measures used in quantitative signal detection has been examined. The different implementations of these measures have not been considered, as these are necessarily dependent on other factors such as the dataset used. For example the MCA use a PRR > 3 and Chi square > 4 and three or more cases as a filter for signal detection.[26] The choice of a suitable method, therefore, also depends on the dataset available. In Table 2 the conditions, advantages and disadvantages in which the various tests can be used are provided. Apart from sensitivity, specificity, positive and negative predictive value, the possibility for correcting for covariates can be useful. For this reason, the Netherlands Pharmacovigilance Foundation Lareb presently uses the ROR − 1.96 SE to calculate disproportionality, since in logistic regression analysis these adjustments can easily be made. Similarly the BCPNN approach can be adapted to adjust for covariates. A major drawback with the ROR, however, is the fact that in the case where an ADR is specifically associated with a certain drug, there is a risk that the number of reports in cell b or c of the contingency table (Figure 1) contain no reports and subsequently the odds ratio cannot be calculated. This situation may occur when rare ADRs are being reported. An example of this type of combination is for instance the detection of a phocomelia associated with thalidomide, or practolol associated with the oculo-mucocutaneous syndrome.

The ROR is a transparent measure, easily interpretable and easily programmed in database programs or spreadsheet programs. An additonal advantage of using the odds ratio is the fact that non-selective underreporting of a drug or adverse drug reaction has no influence on the value of the ROR compared with the population of patients experiencing an ADR.[5]

Disproportionality is simply one way of selecting drug–ADR combinations that may be interesting for clinical review. No individual approach to detect signals is adequate and the concurrent use of other methods is therefore essential.

CONCLUSION

Statistical analyses have been shown to be useful tools in aiding signal detection in spontaneous reporting systems. The various measures that are being applied in quantitative signal detection in various national centres, are comparable when four or more reports constitute the drug–ADR combination. The

Table 2. Conditions, advantages and disadvantages of different measures of disproportionality

Measure of disproportionality	Expected 'null value'	Conditions	Advantage	Disadvantage
ROR – 1.96 SE	1	Cells a,b,c and d have to contain reports	Easy applicable Different adjustments possible in logistic regression analysis In logistic regression analysis, interaction terms can be used for the analysis of drug interactions and syndromes	Odds ratio and standard error cannot be calculated if denominator is zero (specific ADRs) Interpretation difficult Results not always reliable in the event of small numbers in cells a,b,c and d of the contingency table
PRR – 1.96 SE	1	Cells a and c have to contain reports	Easy interpretation	Standard error cannot always be calculated
Yule's Q – 1.96 SE	0	Cells a,b,c and d have to contain reports		Standard error cannot always be calculated Difficult to interpret
IC – 2 SD	0	None	Always applicable Large numbers of calculations can be made efficiently Can be used for pattern recognition in higher dimensions	Relatively non-transparent for people not familiar with Bayesian statistics
Poisson		Only for rare events	Correction for different covariates can be easily established in Poisson regression	Only *p*-value provided
Chi square (Yates correction)			Always applicable	Difficult to interpret

heterogeneity of the data collected in databases of spontaneous reporting systems and the variety of biases influencing data (such as underreporting) are likely to have more influence on the potential for signal detection than the small behavioural differences between the measures detected in this study. Although no 'gold standard' is available, each method has its own advantages and disadvantages in respect to applicability in different situations and possibilities for implementation. Since quantitative signal detection cannot take into account clinical aspects, a case-by-case approach will remain necessary both as an adjunct and an alternative.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Professor I.R. Edwards for his valuable critical comments.

APPENDIX A

(Variables used in the different formulas correspond to the 2×2 contingency table of Figure 1).

Reporting odds ratio (ROR)

The ROR can be expressed as[5]

$$\mathrm{ROR} = \frac{(a/c)}{(b/d)} = \frac{ad}{bc}$$

The standard error of ln(ROR) and 95% confidence interval can be calculated by

$$\mathrm{SE}(\ln \mathrm{ROR}) = \sqrt{\left(\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}\right)}$$

$$95\%\mathrm{CI} = \mathrm{e}^{\ln(\mathrm{ROR}) \pm 1.96 \sqrt{\left(\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}\right)}}$$

Proportional ADR reporting ratio (PRR)

The PRR can be expressed as

$$\text{prr} = \frac{a/(a+b)}{c/(c+d)}$$

The standard error of ln(PRR) and 95% confidence interval can be calculated by[27]

$$\text{SE}(\ln \text{PRR}) = \sqrt{\left(\frac{1}{a} - \frac{1}{a+b} + \frac{1}{c} + \frac{1}{c+d}\right)}$$

$$95\%\text{CI} = \text{e}^{\ln(\text{PRR}) \pm 1.96\sqrt{\left(\frac{1}{a} - \frac{1}{a+b} + \frac{1}{c} + \frac{1}{c+d}\right)}}$$

Poisson probability. The Poisson probability is calculated by[10]

$$p = 1 - \sum_{k=0}^{a-1} \frac{\text{e}^{-\mu} \times \mu^k}{k!}$$

where a is the observed number of reports, and μ is the expected number of reports.

Information component. Resulting from the BCPNN and its variance can be calculated as[17]

$$E(IC_{ij}) = \log_2 \frac{(c_{ij} + \gamma_{ij})(C + \alpha)(C + \beta)}{(C + \gamma)(c_i + \alpha_i)(c_j + \beta_j)}$$

$$V(IC_{ij}) = \frac{\dfrac{C - c_{ij} + \gamma - \gamma_{ij}}{(c_{ij} + \gamma_{ij})(1 + C + \gamma)} + \dfrac{C - c_i + \alpha - \alpha_i}{(c_i + \alpha_i)(1 + C + \alpha)} + \dfrac{C - c_j + \beta - \beta_i}{(c_i + \beta_j)(1 + C + \beta)}}{(\log 2)^2}$$

where

$$\gamma = \gamma_{ij} \frac{(C + \alpha)}{(c_i + \alpha_i)} \cdot \frac{(C + \beta)}{(c_j + \beta_j)}$$

and $\gamma_{ij} = 1,\ \alpha_i = 1,\ \alpha = 2, \beta_j = 1, \beta = 2,\ C$ is the total number of reports in the database, C_{ij} the number of combinations between a specific drug [i] and the suspected adverse drug reaction [j], C_i the total number of reports on drugs [i] in the database and C_j the total number of reports on the suspected ADR [j] in the database.

Chi square (Yates' correction)

Chi square tests for a 2×2 table, with Yates' correction can be expressed as

$$\chi^2 = \sum \frac{(|O - E|)^2 - \frac{1}{2}}{E}$$

The summation applies over all four cells of the contingency table. O is the observed frequency and E is the expected frequency of the reports. For example, in case the contingency table is used for the first cell O and E should be calculated as:

$$O = a \qquad E = \frac{(a+b)(a+c)}{(a+b+c+d)}$$

For the other cells it takes the value contained in the cell, i.e. b, c, d in turn

Yule's Q. Yule's Q can be expressed as[28]

$$Q = \frac{ad - bc}{ad + bc}$$

The standard error of Yule's Q and the 95% CI is calculated by

$$\text{SE}_Q = \frac{1}{2}\left(1 - Q^2\right)\sqrt{\left(\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}\right)}$$

$$95\%\,\text{CI} = Q \pm 1.96\,\text{SE}_Q$$

REFERENCES

1. Olsson S. The role of the WHO programme on International Drug Monitoring in coordinating worldwide drug safety efforts. *Drug Saf* 1998; **19**: 1–10.
2. Finney DJ. The design and logic of a monitor of drug use. *J Chron Dis* 1964; **18**: 77–98.
3. Edwards IR, Lindquist M, Wiholm B-E, Napke E. Quality criteria for early signals of possible adverse drug reactions. *Lancet* 1990; **336**: 156–158.
4. Finney DJ. Statisical logic in the monitoring of reactions to therapeutic drugs. *Methods Inf Med* 1971; **10**: 237–245.
5. Stricker BHCh, Tijssen JGP. Serum sickness-like reactions to cefaclor. *J Clin Epidemiol* 1992; **45**: 1177–1184.
6. Evans SJW, Waller P, Davis S. Proportional Reporting Ratios: The uses of epidemiological methods for signal generation. *Pharmacoepidemiol Drug Safe* 1998; **7**(Suppl. 2): S102 (abstract no. 55).
7. Egberts ACG, Van der Hofstede JW, Meyboom RHB, *et al.* Egberts ACG. Pharmacoepidemiologic approaches to the evaluation of antidepressant drugs (thesis). Utrecht: 1997;

4.1, Transformation of a database of spontaneously reported suspected adverse drug reactions and its use as a tool in signal detection. 111–124.
8. Evans SJW. Pharmacovigilance: a science or fielding emergencies? *Stat Med* 2000; **19**: 3199–3209.
9. Moore N, Kreft-Jais C, Haramburu F, *et al.* Reports of hypoglycaemia associated with the use of ACE inhibitors and other drugs: a case/non-case study in the French pharmacovigilance system database. *Br J Clin Pharmacol* 1997; **44**: 513–518.
10. Tubert P, Begaud B, Pere JC, Haramburu F, Lellouch J. Power and weakness of spontaneous reporting: a probabilistic approach. *J Clin Epidemiol* 1992; **45**: 283–286.
11. Hayes WL. The Poisson distribution. In *Statistics* (4 edn). Holt, Rinehart and Winston, Inc.: Fort Worth, 1988; 144–146.
12. 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**: 315–321.
13. Lindquist M, Ståhl M, Bate A, Edwards IR, Fucik H, Nunes AM. From association to alert—a revised approach to international signal analysis. *Pharmacoepidemiol Drug Safe* 1999; **8**: S15–S25.
14. Lindquist M, Ståhl M, Bate A, Edwards IR, Meyboom RH. 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**: 533–542.
15. DuMouchel W. Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. *Am Stat* 1999; **53**: 177–189.
16. Broekmans AW, Lekkerkerker JFF, De Koning GHP, Vree PH. Nieuwe regels voor het melden van bijwerkingen in Nederland na 1995. *Ned Tijdschr Geneeskd* 1996; **140**: 1166–1167.
17. Orre R, Lansner A, Bate A, Lindquist M. Bayesian neural networks with confidence estimations applied to data mining. *Comput Stat Data Anal* 2000; **34**: 473–493.
18. Waller P, Lee EH. Responding to drug safety issues. *Pharmacoepidemiol Drug Safe* 1999; **8**: 535–552.
19. Meyboom RHB, Egberts ACG, Edwards IR, Hekster YA, de Koning GHP, Gribnau FW. Principles of signal detection in pharmacovigilance. *Drug Saf* 2000; **16**: 355–365.
20. Hayes WL. The assumptions in chi-square tests of association. In *Statistics* (4 edn). Holt, Rinehart and Winston, Inc.: Fort Worth, 1988; 780–781.
21. Tubert-Bitter P, Bégaud B. Comparing safety of drugs. *Post Marketing Surveill* 1993; **7**: 119–137.
22. Bègaud B, Moride Y, Tubert-Bitter P, Chaslerie A, Haramburu F. False-positives in spontaneous reporting: should we worry about them? *Br J Clin Pharmacol* 1994; **38**: 401–404.
23. Amery WK. Assessment of signals generated form spontaneously reported adverse events. *Pharmacoepidemiol Drug Safe* 1999; **8**: 301–304.
24. Meyboom RH, Egberts AC, Gribnau FW, Hekster YA. Pharmacovigilance in perspective. *Drug Saf* 1999; **21**: 429–447.
25. Amery WK. Signal generation from spontaneous event reports. *Pharmacoepidemiol Drug Safe* 1999; **8**: 147–150.
26. Evans SJW. Recognition of signals from spontaneous reporting systems—Comparison of different approaches: Poisson model, neural networks, proportional reporting ratios. In *Proceedings of 20th Annual Conference of the International Society for Clinical Biostatistics* 1999; **119**: (abstract).
27. Greenland S, Rothman KJ. Introduction to categorical statistics. Greenland S, Rothman KJ (eds). In *Modern Epidemiology* (2nd edn), Lippincott-Raven: Philadelphia, 2001; 231–252.
28. Kendall M, Stuart A. *The Advanced Theory of Statistics.* Griffin (Publ): London, 1995.

