Drug Saf (2013) 36:687–692
DOI 10.1007/s40264-013-0075-1

ORIGINAL RESEARCH ARTICLE

Choosing Thresholds for Statistical Signal Detection with the Proportional Reporting Ratio

Jim Slattery · Yolanda Alvarez · Ana Hidalgo

Published online: 11 June 2013

Abstract
Background Identification of potential drug safety problems using statistical screening algorithms in routinely collected databases of adverse drug reactions (ADRs) requires decision rules based on thresholds of the chosen parameters. Choosing higher or lower thresholds changes both the sensitivity of the screening and the number of false alarms produced, and thus has an impact on the effectiveness of the detection process.
Objective The aim of this study was to evaluate the impact on the effectiveness of signal detection activities of choosing different warning thresholds for the proportional reporting ratio (PRR) and for the count of reports of any drug-event combination.
Methods Signal detection methods were tested within the EudraVigilance database of suspected ADRs. Using an established set of known ADRs, the number that could be detected and the changes in time gained for earlier investigation of the signal were calculated over a range of signal detection thresholds. These figures were set against the number of false positive signals produced by the statistical signal detection algorithms.
Results Higher thresholds for the lower confidence bound of the PRR produced fewer false positives but this benefit was offset by important losses of sensitivity in the detection of ADRs. By contrast, increases in the threshold for the count of a specific drug-event combination produced fewer false positives with little loss of either sensitivity or time gained for investigation of adverse events. A threshold of five compared with the current European Medicines Agency threshold of three gave a reduction of 25 % in false positive signals in return for a loss of 12 % in true signals detected early.
Conclusion Changes in the standard threshold for the count of drug-event combinations can result in a substantial improvement in efficiency of the signal detection process. Initially this change might be applied only to products with a well-established safety profile.

J. Slattery (✉) · Y. Alvarez · A. Hidalgo
Pharmacovigilance and Risk Management Sector,
European Medicines Agency, 7 Westferry Circus, Canary Wharf,
London E14 4HB, UK
e-mail: Jim.Slattery@ema.europa.eu

1 Introduction

Statistical methods for detection of adverse drug reactions (ADRs) to medicines are commonly based on exploratory analysis of databases of clinical information. Most pharmaceutical companies and regulatory authorities compile such databases from reports of possible ADRs that name the suspect medicine and describe the nature of the clinical event. The decision to investigate further a potential adverse reaction is then based on a calculated ratio of the actual number of such adverse events reported with a given drug to the number that would be expected based on the assumption that the chance of a report of this event with the drug was the same as the chance of such a report with any other drug. Such a measure is referred to as a disproportionality statistic (DS) [1]. If this ratio becomes large, suspicions are raised and further action initiated. The precise threshold at which a DS triggers in-depth evaluation is based on empirical considerations. Not least the past experience of the positive predictive value of such tests and the resources available to perform the more labour-intensive exploration of the possible problem.

In addition to considering the value of the DS it is common to place some lower limit on the absolute number

of reports for a particular drug/adverse event combination. This is because it is known that small counts of rare events may give somewhat erratic signal detection based on the DS alone.

In 2010 the European Medicines Agency (EMA) performed an evaluation of the signal detection performance in EudraVigilance based on the established signalling criteria [2]. These are a formal lower 97.5 % confidence bound of 1 for the proportional reporting ratio (PRR025 >1) and a count of 3 or more reports related to the specific problem. These thresholds are quite widely used, not only at the EMA but in other signal detection systems. The analysis presented here is a revision of these study results at a range of values of both thresholds—that for PRR025 and that for the simple count of drug/adverse event pairs. The question asked is whether the EMA EudraVigilance data support the use of these thresholds.

The primary benefit of any signal detection activity is the additional time to assess emerging safety issues gained through earlier detection and the opportunity to make earlier interventions to protect public health. This benefit is weighed against the number of additional signals of disproportionate reporting (SDRs) that require evaluation when statistical screening is introduced. The evaluation of different thresholds in this study uses outcome measures that reflect these potential benefits and drawbacks of statistical screening methods.

2 Methods

The evaluation of signalling performance was conducted in two distinct parts. The first of these assessed the effectiveness of the statistical technique in finding known safety issues. The second part assessed the numbers of additional potential safety signals that were generated by the statistical procedure and hence the amount of work required to operate the system. During the current EMA signal detection activities, any identified signal subsequently undergoes a standard evaluation procedure that may result in a regulatory action or, more often, in the conclusion that a false positive has been observed. Sometimes such evaluations are simple and sometimes they involve considerable effort; thus, the costs and consequences of the positives generated are not easy to specify, but the number of positives is a reasonable surrogate measure of the workload penalty resulting from a signal detection system. In the previous paper of Alvarez et al. [2], the list of signals detected was subjected to an initial triage to eliminate those that were easily explicable through known confounding mechanisms. Due to the large number of different signalling criteria used in this study, we have omitted this step and hence the results are not directly comparable to the previous study. However, the comparison between alternative signalling criteria is not compromised by omission of this step.

A set of 267 centrally authorized products (CAPs) for marketing across the EU [3] was selected for study on the basis that the marketing authorization holder had been sending electronic reports to the EudraVigilance database for at least 1 year by the start of the study time window. CAPs were chosen because the EMA has access to full documentation of regulatory activities for these products. All CAPs that fulfilled the selection requirement were used. Reports used were the same subset that is used in routine EMA signal detection activities. The set of ADRs against which the statistical signal detection methods were assessed consisted of adverse events that were added to the Summary of Product Characteristics (SPC) for the 267 products within a time window between September 2003 and March 2007. All ADRs that were Preferred Terms (PTs) of the Medical Dictionary for Regulatory Affairs (MedDRA®)[1] classification [4] included on the EMA list of Important Medical Events (IMEs) were included. This list is available on request from the EudraVigilance website [5].

This set of known safety issues was investigated over a time period extended back to product authorization, and possibly much before September 2003, in order to classify them as either predicted or not predicted by statistical signal detection. The total number of positive signals—both true and false—arising for the 267 products within the study time window was then calculated as a measure of the workload associated with the signal detection process. This method allowed the study to cover a wide range of products at all stages of the product lifecycle.

3 Statistical Methods

The comparison between different signal detection thresholds is made in terms of three parameters: firstly, the ratio of the total number of signals arising to the number of ADRs that would be detected earlier; secondly, the proportion of ADRs detected earlier; and thirdly, the median time gained for each true signal detected early.

The PRR is an estimate of the probability that a spontaneous report containing a product (X) will mention an adverse event (Y) divided by the probability that a report not containing X will mention Y. The calculation and use of the PRR is discussed by Evans et al. [6], and details of

[1] MedDRA® terminology is the international medical terminology developed under the auspices of the International Conference on Harmonization (ICH) of Technical Requirements for Registration of Pharmaceuticals for Human Use.

the way it is used in signal detection at the EMA can be found in the EMA guidelines [7]. Use of the PRR in this study reflected current practice at the EMA and, in particular, no adjustment was made for possible confounding variables.

4 Assessment of Effectiveness of Proportional Reporting Ratio (PRR) Signal Detection

The agreed profile of adverse reactions of a medicine authorized in the EU is contained in an SPC. The SPC is updated through regulatory procedures called type II variations. The EMA database for tracking regulatory procedures (SIAMED, WHO© 1998) was searched to identify all type II variations to the marketing authorization for the study products over the period of the study time window, and all safety issues added to the SPC (section 4.4 or 4.8) were identified by direct inspection. A researcher then reviewed EMA documentation to find the earliest date at which the EMA became aware that a safety signal requiring investigation had been identified. This date was adopted as the original date of the discovery of the signal from established pharmacovigilance activities (index date). The safety issue was then mapped to a MedDRA® PT or group of PTs. As in our previous work [2], only IMEs were included and clear synonyms were identified for some of the listed terms; a signal relating to any synonymous term was considered a successful identification of the ADR.

The safety issues identified in the scope of the type II variations were divided into two classes for each SDR criterion: (1) those for which either no SDR was observed in EudraVigilance or for which the SDR arose after the index date; and (2) those for which the SDR preceded the index date. Only these latter reflect an advantage for statistical signal detection in that they allow the signal validation process to start earlier.

5 Assessment of False Positives Associated with PRR Signal Detection

During the study period, for each product a number of adverse events, other than those which were in the scope of the variations, would be expected to reach the thresholds that define an SDR. For the analysis, all such SDRs were treated as false positives. Many such SDRs result from confounding (typically, the event is linked to the underlying disease for which the product is prescribed rather than being a true adverse effect of the medicinal product) or reflect the stochastic nature of the process by which adverse events arise. An ad hoc routine was written in Statistical Analysis Systems, Version 9.1 (SAS Software, Cary, NC, USA) computer code to calculate the PRRs as a function of time for every product in the study and to locate those that reached, for the first time, the defined threshold for an SDR within the appropriate time window for the study. The effort involved in assessment of these SDRs can be considered as the cost to be paid (from a regulatory resource point of view) for the use of an additional signal detection procedure. From a public health perspective it is also important that these false positives can adversely affect availability of effective medicines while they are being investigated.

6 Results

The number of SDRs raised per ADR predicted earlier than by established pharmacovigilance methods is shown in Fig. 1, and the proportion of true effects detected early by an SDR is shown in Fig. 2. Ideally we would choose thresholds that give low values in Fig. 1 but high values in Fig. 2.

Of equal importance to the number of ADRs for which we can receive early warning through the statistical signal detection methods is the amount of early warning. Figure 3 shows the time in months by which the SDR preceded the first warning received via other methods of pharmacovigilance. This quantity determines the time gained for evaluation of the signal and any subsequent regulatory action. Figure 3 shows how this time gained varies as the threshold for the number of reports varies. The threshold for PRR025 is held constant at 1 for this calculation.

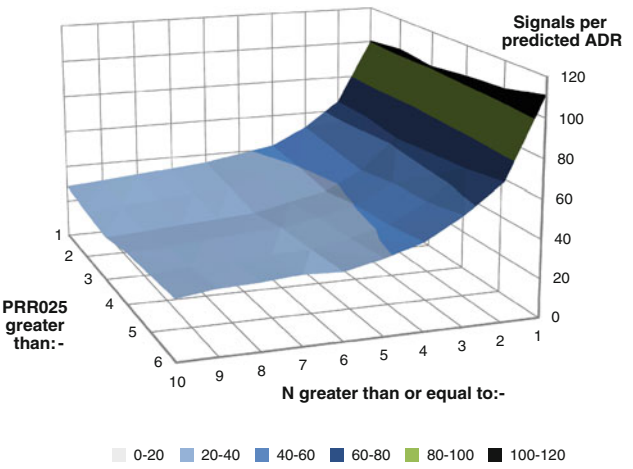

Fig. 1 Number of signals per successful early prediction. *ADR* adverse drug reaction, *PRR* proportional reporting ratio

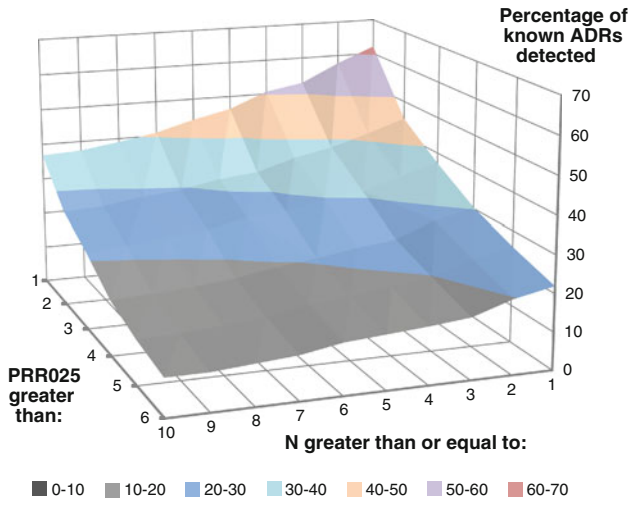

Fig. 2 Percentage of ADRs pre-empted. *ADRs* adverse drug reactions, *PRR* proportional reporting ratio

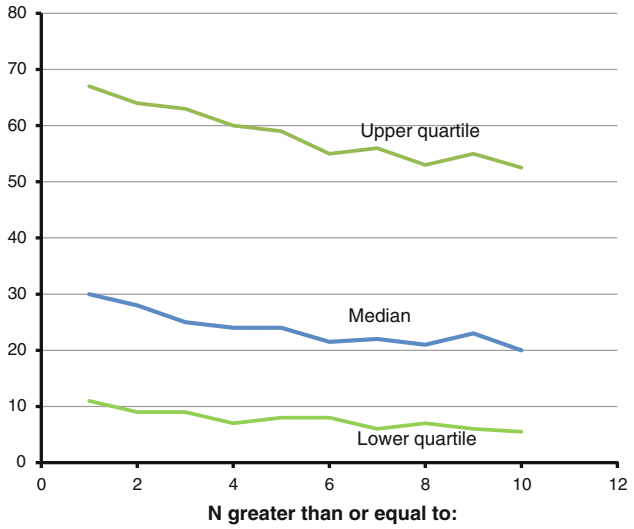

Fig. 3 Median time gained for each signal pre-empted (months)

7 Discussion

The changes in outcome measures with changes in the thresholds for the number of reports (*N*) and for PRR025 are very clear in form and of a fairly large magnitude. Hence the interpretation is straightforward. Perhaps the most surprising finding from Fig. 1 is that raising the threshold for PRR025 does not improve the ratio of SDRs to true signals. The graph is generally flat in this direction. For instance, at the conventional $N \geq 3$, a threshold for PRR025 of 1 produces 51.6 positive signals for each true ADR detected, and the value for a threshold of 5 is 51.5. By contrast, raising the threshold for *N* initially markedly increases the efficiency of the statistical signal detection process but this advantage is mostly realised by the time that five or six reports have been received. After that the graph is relatively flat and only minor changes in efficiency are obtained. At the currently used PRR025 >1, an increase in the threshold for *N* from 3 to 5 reduces the positive signals per ADR from 51.6 to 38.9, a reduction of 25 %.

Raising either threshold resulted in some of the known ADRs not being detected early (see Fig. 2). However, this effect is less marked with changes in the threshold for *N* than in the threshold for PRR025. At PRR025 >1, a change in the *N* threshold from the current value of 3 to a value of 5 resulted in a reduction in the proportion of ADRs detected from 52 to 46 %. At $N \geq 3$, a change in the PRR025 threshold from 1 to 2 reduced the proportion detected from 52 to 38 %.

From the above results it appears obvious that no increase in the PRR025 threshold should be adopted but that a higher value for the threshold in *N* could be warranted. However, waiting for a larger number of reports to arrive in EudraVigilance must reduce the advantage in time gained. Thus, the remaining question is what time penalty is paid for an increase in this threshold. Figure 3 shows that the median time gain fell from 25 months at the current value of three or more reports to 24 months at five or more reports. At six reports this was estimated to be 21.5 months. This rather small loss may appear surprising but, of course, it is averaged over the ADRs detected early and these have reduced in number somewhat as the threshold increases. It is also likely that those lost will often correspond to those with a small time advantage. In fact, the median time gain in the 23 signals lost on increasing the threshold from 3 to 5 is only 7 months. It is also worth noting that the random fluctuation in Fig. 3 appears to favour the estimate at $N = 5$, whilst lowering that at $N = 6$.

In this study we addressed a single aspect of statistical signal detection and considered only one DS in a single large reporting database. Whether these results impact on signal detection using other methodologies is unclear. Current work to examine multifactorial aspects of signal detection and other DSs is ongoing, not least in the Innovative Medicines Initiative PROTECT (Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium) project led by EMA [8]. Hence we hope to be able to address these shortcomings in the near future.

Other researchers have investigated statistical signal detection thresholds. Matsushita et al. specifically addressed use in pharmaceutical company databases [9]. This study found low sensitivities for current standard criteria and recommended criteria that generated more positive signals. This argument holds weight if statistical signal detection is considered as an initial filter for selection of potential ADRs. This is not the way that statistical signal detection is used at

the EMA, where it is seen as just one component of a process employing multiple clinical and statistical approaches in parallel and the effect of changes in sensitivity for one component is not so easy to evaluate. Almenoff et al. [10] discussed briefly adjustment of the lower bound on the Empirical Bayes Geometric Mean (EB05) and PRR signalling thresholds to equalize the numbers of SDRs and hence facilitate comparison of the two methods. Ahmed et al. [11] compared three established signal detection methods with two false discovery rate (FDR) methods. The metrics for the five-way comparison were the number of drug-event combinations detected from a reference set investigated by the French regulatory agency and the time to detection. In this study, two different thresholds of FDR were examined but the thresholds for the methods already in routine use, including PRR, were kept constant and based on the formal lower confidence bound without restriction on the number of reports. In terms of the chosen metrics, Ahmed et al. [11] found advantages for the FDR methods but the absence of a measure of total workload generated by the conventional methods leaves doubt concerning the relative impact of the methods on the functioning of a pharmacovigilance system. Berlin et al. [12] assessed a number of different signalling thresholds for EB05 and an odds ratio from logistic regression (LR0005 and LR05). The reference standard used for ADRs was the company core data sheet and results were expressed in terms of sensitivity, specificity and positive predictive value. A fixed threshold of 4 was used for the number of cases.

The applicability of the results of this study and the optimal threshold for *N* needs some thought. In a pharmacovigilance department with low workloads, any reduction in sensitivity may not be acceptable. However, experience suggests that most departments, despite working at full capacity, are forced to triage potential signals to prioritize them for assessment. Under these circumstances, a reduction in false signals per ADR will increase the number of new ADRs discovered because the number of assessments performed will remain constant. Notwithstanding this, the introduction of new signalling criteria should be exercised with care. Application to selected products or subsets of ADRs may be considered. For example, changes in the signal detection thresholds could be applied first to products with longer established safety profiles, allowing more intensive effort to be focused on new and less well-characterized drugs. On the basis of this study, a threshold of either 4, 5 or 6 would improve efficiency compared with our current standard of 3. A conservative change to 4 would realise much of the available improvement; on the other hand, the argument above suggests that busy departments might benefit from increasing to 6. A rational approach appears to be that a threshold of 5 be adopted with ongoing prospective evaluation.

8 Conclusion

The message from this re-analysis of our previous study appears to be that a small increase in the threshold for *N* from 3 to 5 will substantially reduce the proportion of SDRs identified in the EudraVigilance database of suspected ADRs that prove to be false. The penalty paid for this is that some ADRs can no longer be detected through statistical signal detection, but these will still be detected by other means and only small losses in the time available for evaluation will be incurred. Moreover, in a resource-limited system, the lost ADRs will be compensated by the opportunity to evaluate other true SDRs. The total balance should result in a reduction of the workload per ADR detected with statistical signal detection using the PRR which, with no increase in workload, will allow a larger number of ADRs to be identified.

Acknowledgments The authors would like to thank Prof. H.-G. Eichler, Prof. M. Posch and F. Maignen for their constructive comments. The views expressed in this article are the personal views of the author(s) and may not be understood or quoted as being made on behalf of or reflecting the position of the EMA or one of its committees or working parties.

Funding No sources of funding were used to conduct this study or prepare this manuscript.

Conflict of interest Jim Slattery, Yolanda Alvarez and Ana Hidalgo have no conflicts of interest to declare that are directly relevant to the content of this study.

References

1. Finney DJ. Statistical logic in the monitoring of adverse reactions to therapeutic drugs. Methods Inf Med. 1971;10:237–42.
2. Alvarez Y, Hidalgo A, Maignen F, Slattery J. Validation of statistical signal detection procedures in EudraVigilance post-authorisation data: a retrospective evaluation of the potential for earlier signalling. Drug Saf. 2010;33(6):475–87.
3. Council Regulation (EEC) No 2309/93 of 22 July 1993 laying down Community procedures for the authorization and supervision of medicinal products for human and veterinary use and establishing a European Agency for the Evaluation of Medicinal Products. http://ec.europa.eu/health/files/eudralex/vol-1/reg_1993_2309/reg_1993_2309_en.pdf. Accessed 16 May 2013.
4. MedDRA and MSSO. http://www.meddramsso.com. Accessed 18 Dec 2012.
5. European Medicines Agency. EudraVigilance. http://eudravigilance.emea.europa.eu/highres.htm. Accessed 18 Dec 2012.
6. Evans SJW, Waller P, Davis S. Use of proportional reporting ratios (PRR) for signal generation from spontaneous adverse drug reaction reports. Pharmacoepidemiol Drug Saf. 2001;10:483–6.
7. EudraVigilance Expert Working Group. Guideline on the use of statistical signal detection methods in the EudraVigilance data analysis system (EMEA/106464/2006 rev.1). http://www.emea.europa.eu/pdfs/human/phvwp/10646406enfin.pdf. Accessed 18 Dec 2012.

8. Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium. http://www.imi-protect.eu/. Accessed 18 Dec 2012.
9. Matsushita Y, Kuroda Y, Niwa S, Sonehara S, Hamada C, Yoshimura I. Criteria revision and performance comparison of three methods of signal detection applied to the spontaneous reporting database of a pharmaceutical manufacturer. Drug Saf. 2007;30(8):715–26.
10. Almenoff J, LaCroix K, Yuen N, et al. Comparative performance of two quantitative safety signalling methods. Drug Saf. 2006; 29(10):875–87.
11. Ahmed I, Thiessard F, Miremont-Salame G, et al. Early detection of pharmacovigilance signals with automated methods based on false discovery rates. Drug Saf. 2012;35(6):495–506.
12. Berlin C, Blanch C, Lewis D, et al. Are all quantitative post-marketing signal detection methods equal? Performance characteristics of logistic regression and multi-item Gamma Poisson Shrinker. Pharmacoepidemiol Drug Saf. 2012;21:622–30.