ORIGINAL RESEARCH ARTICLE

Drug Saf 2010; 33 (6): 475-487
0114-5916/10/0006-0475/$49.95/0

Validation of Statistical Signal Detection Procedures in EudraVigilance Post-Authorization Data

A Retrospective Evaluation of the Potential for Earlier Signalling

Yolanda Alvarez,[1,2] *Ana Hidalgo*,[1] *Francois Maignen*[1] and *Jim Slattery*[1]

1 European Medicines Agency, London, UK
2 Department of Pharmacology, Therapeutics and Toxicology, Universitat Autonòma de Barcelona, Barcelona, Spain

Abstract

Background: Screening large databases of spontaneous case reports of possible adverse drug reactions (ADRs) is an established method of identifying hitherto unknown adverse effects of medicinal products; however, there is a lack of consensus concerning the value of formal statistical screening procedures in guiding such a process. This study was performed to clarify the nature of any added benefits and additional effort required when established pharmacovigilance techniques are supplemented with statistical screening.

Objective: To evaluate whether statistical signal detection in spontaneous reporting data can lead to earlier detection of drug safety problems and to assess the additional regulatory work entailed.

Methods: Using the EudraVigilance post-authorization module (EVPM), a screening procedure based on the proportional reporting ratio (PRR) was applied retrospectively to examine if regulatory investigations concerning ADRs in a predefined set of products could have been initiated earlier than occurred in practice. During the same time period, between September 2003 and March 2007, the number of PRR-based signals of disproportionate reporting (SDR) that arose in the same set of products was calculated and evaluated to determine the number requiring investigation. The outcome is expressed as the ratio of the number of SDRs requiring investigation compared with the number of signals pre-empted by the statistical screening approach. In those cases where the signal was discovered earlier, the delay was calculated between identification by the PRR method and by the method that originally identified the signal.

Results: In 191 chemically different products, 532 adverse reactions were added to the summary of product characteristics during the study period. Of these, 405 were designated as important medical events (IMEs) based on a comprehensive predefined list. Of the IMEs, 217 (53.6%) were identified earlier by the statistical screening technique, 79 (19.6%) were detected after

the date at which they were raised by standard pharmacovigilance methods and 109 (26.9%) were not signalled during the study period. 1561 SDRs requiring further evaluation were detected during the study period, giving a ratio of 7.2 assessments for each signal pre-empted. The mean delay between the discovery of signals using the statistical methods in the EVPM and established methods in the 217 cases detected earlier was 2.45 years. A review resulted in clear explanation for why the statistical method had not pre-empted detection in all but 77 of 188 cases.

Conclusions: The form of statistical signal detection tested in this study can provide significant early warning in a large proportion of drug safety problems; however, it cannot detect all safety issues more quickly than other pharmacovigilance processes and hence it should be used in addition to, rather than as an alternative to, established methods.

Background

For a number of years, one of the principal methods used to identify adverse drug reactions (ADRs) during the post-authorization period has been inspection and analysis of spontaneous reports from health professionals submitted under various reporting schemes. The national reporting schemes in Europe generally ask that a report be submitted when an adverse event has occurred that the sender suspects may be associated with one of the medicines being taken by the patient affected. The intended circumstances under which reports should be submitted and the information that they should contain are precisely specified in EU guidelines[1,2] and codified in the data collection forms. Thus, there is a homogeneity to the data collection which suggests that combination of data from such reports may induce a certain synergy. In other words, that the combined data may yield more than consideration of the individual case reports. This thought has led to a number of statistical procedures being suggested as promising ways of extracting information about drug safety from the datasets: so-called signal detection methods.[3]

Evaluation of signal detection methods and comparison with alternative forms of pharmacovigilance in the setting of real-life drug safety problems has proved a challenge. The concern that a group of patients is currently exposed to the product dictates that drug safety investigations are swiftly undertaken on the basis of all available information combined; thus, reliably contrasting the impact of different sources of information is difficult. However, as a result of backlog reporting, an opportunity to make a relatively independent evaluation of signal detection in a spontaneous reporting database and other methods of pharmacovigilance has recently arisen in the EudraVigilance database.

The EudraVigilance[4] post-authorization module (EVPM) is a database intended to be a single repository for all reports of possible serious adverse reactions that concern medicines subject to a marketing authorization within the EU. It was set up by the European Medicines Agency (EMA) in collaboration with EU Member States, and the first electronic transmissions were in December 2001. Reports are sent by EU regulatory agencies and pharmaceutical companies. A period of some years was required for all reporters to meet the technical conditions for sending electronic reports and this only became a legal requirement in November 2005. The EVPM dataset will be complete when the backlog of reports reported between the creation of the EU pharmacovigilance system[1] on 1 January 1995 and the start of electronic reporting by each sender has been transmitted into the EVPM. This process is now in its final stages, with more than 500 000 backlog reports received. The detection

of new safety signals relies on the current dataset of more than 1 million individual case safety reports, transmitted mostly since November 2005.

The backlog reporting is the feature of EVPM, which made it different from other spontaneous reporting databases for much of the setting up period. Although it now contains almost all the data that it would have contained had it been set up using a prospective reporting system since 1995, it in fact represents a compilation of data that was not available to the EMA during this time. Hence, it could not have a direct influence on regulatory decisions during the period of backlog reporting. Thus, comparison of retrospective analyses of EVPM with historical details of actual regulatory activities can give information about the way in which these activities might have been changed if the EVPM data had been fully available at the time.

Methods

The study 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.

A set of 267 centrally authorized products for marketing (CAPs) across the EU[5] was selected for study on the basis that the marketing authorization holder had been sending electronic reports to the EVPM for at least 1 year. Although the system was still in a preliminary phase at this time, it was intended by this criterion to select products for which the signal detection capabilities of the system should be similar to those of a fully established system. The criterion did not differentiate between new and old products and hence data were collected over a wide range of times from authorization. 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. Not all EudraVigilance reports were used in this study, the main exclusion being literature reports, but those used were the same set that is used in routine EMA signal detection activities.

The study design uses elements of both cross-sectional and cohort methods. Initially, the known safety issues and potential safety signals were assessed within a time window between September 2003 and March 2007. Thus, this comparison can be seen as based on a cross-sectional viewpoint, albeit with an extended temporal dimension. However, the group of known safety issues were then 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; hence, this latter was more akin to a cohort study. This combination of methods allowed the study to cover a wide range of products at all stages of the product life-cycle.

Statistical Methods

The primary outcome of the study is the additional time to assess emerging safety issues gained through earlier detection, and this is weighed against the number of additional signals of disproportionate reporting (SDRs) that require evaluation when statistical screening is introduced. The outcomes were chosen to reflect the potential practical benefits and drawbacks of statistical screening methods. Despite some analogies between ADR detection and screening in clinical settings, the traditional measures of sensitivity and specificity cannot be satisfactorily estimated as only a partial set of ADRs is known; thus, there is no equivalent of a clinical gold-standard test. The distribution of delays

between SDRs and signals from other pharmacovigilance methods is presented as Kaplan-Meier curves and confidence intervals (CIs) for statistics based on such curves using standard techniques. CIs on proportions assume binomial distributions.

The proportional reporting ratio (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 the way it is used in signal detection at the EMA can be found in EMA guidelines.[7] No adjustment was made in this study for possible confounding variables. The rule used to define an SDR is that the lower bound of the central 95% CI on the PRR is >1, and three or more reports have been received naming the relevant product and adverse event.

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 a summary of product characteristics (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, 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 Medical Dictionary for Regulatory Affairs (MedDRA®)[1] preferred term (PT) or group of PTs.[8]

Although the EU guideline[1] on SPCs requires the use of MedDRA (preferably at the PT level) to describe adverse reactions in the undesirable effects section (section 4.8) of the SPC, effecting a map between the MedDRA terms used in the SPC and those used in ADR reporting is not always easy. In part, the difficulty arises because different MedDRA PT codes may be used in each setting to describe equivalent adverse events. There are two likely reasons for this. First, as the MedDRA is designed to allow accurate coding of the precise term reported by a medical professional, there are numerous groups of almost synonymous terms, and the choice of which to use may simply depend on which term is located first. Second, regulatory coding is done on the basis of the best evidence available and attention is given to choice of the most specific term, whilst coding for adverse event reporting is done on the basis of limited clinical data and is likely to employ less precise terms. The result of this is that searching for a PRR signal based only on the exact term used in a variation may not reveal the true performance of the system. In order to handle this problem in the context of this study we decided to do the analysis both on the basis of exact matches alone and on the basis that a signal detected from any of a small group of synonymous PTs might equally be considered to have had the potential to initiate a regulatory action equivalent to that under study. With regard to the calculation of disproportionality statistics for the group of synonyms, it is important to understand that the calculations were done separately for each synonym rather than combining the synonyms into a group. Thus, the process reflected the procedure followed in standard signal detection activities and did not create an *ad hoc* intermediate level between the PTs and higher level terms in the MedDRA hierarchy. It seems likely that the creation of such a layer to eliminate synonyms might indeed improve the effectiveness

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.

of signal detection but would not be appropriate for this study, which seeks to evaluate the effectiveness of signal detection based on the current guidelines.

A statistical difficulty with this procedure is that the synonyms had to be defined in a *post hoc* fashion. This definition was, of course, done without knowledge of whether a PRR signal would arise; however, there is a possibility of bias favouring the PRR method as the probability of a signal may rise as synonyms are added to the group. For this reason we did our main analysis based on the groups of synonyms (which is a more realistic reflection of practical signal detection), but also carried out an analysis based on exact MedDRA matches – which should be regarded as a pessimistic assessment but demonstrably not biased by *post hoc* selection of MedDRA terms.

Using the individual MedDRA codes, the PRRs and CIs were calculated as a function of time over the study period. From these calculations it was possible to determine the date at which the first SDR arose within the EVPM dataset for each issue. For this calculation, reports sent to EVPM as 'backlog' were counted as though they had arrived in the timeframe specified by the regulatory requirement for the fully established system.

Use of the MedDRA PT as the principal coding level for signal detection has been chosen on the basis that this level of the MedDRA hierarchy is commonly used in many other signal detection systems. This level currently includes around 18 200 terms that are of variable clinical relevance for signal detection procedures. In order to exclude irrelevant terms we applied a pre-specified filter for seriousness, which left 9078 PTs designated as important medical events (IMEs). The effect and importance of this filter will be discussed in the results section. The list is available on request from the EudraVigilance website.[4]

The safety issues identified in the scope of the type II variations were divided into two classes: (i) those for which either no SDR was observed in EudraVigilance or for which the SDR arose after the index date; and (ii) those for which the SDR preceded the index date.

Assessment of False Positives to Reflect Workload 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. 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 (SAS® 9.1, 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. An evaluation of these SDRs was carried out to determine the regulatory resources entailed in adding the PRR signal detection process to the established pharmacovigilance procedures. The list of SDRs identified was then reviewed and classified by internal assessors. The classification system used was based on that published by Hochberg et al.[9]

Results

Description of Study Products

The 267 medicinal product trade names selected for the study corresponded to 191 chemically different products (i.e with a unique combination of active substances).[2] The counting of ADRs and all analysis was performed within categories

2 This is explained by the fact that some Marketing Authorization Holders have submitted several applications (under different invented names) for the same molecule or active substance (i.e. multiple applications).

defined by active substances as this avoids repeated counting of the same safety issues. In the following discussion, the word 'product' refers to the unique active substance combinations. The type II variations that came into effect within the study window affected 102 (53%) of the active substance combinations and added a total of 532 ADRs to the SPCs.

The types of products in the study were typical of the CAPs, with 14% for treating infections, mainly HIV, 10% for malignant disease, 7% classified as for the CNS and 6% for the endocrine system, while other pharmaceutical classifications each represented <3% of the total. The products were manufactured by 50 different organizations and represented a mixture of general practice and hospital-based treatments. We estimate that about 91 of the 191 products would be routinely used in general practice, while 72 were almost exclusively for use in specialist care; these were predominantly delivered via an intravenous route but also included products aimed at conditions for which the patient would be admitted because of the severity of disease, necessity of other forms of care or possibility of adverse reactions. The remaining 28 tended to be second-line or for rare conditions, and the location of use was less clear. This evaluation was only done for study products but the overall proportions of products that could be delivered intravenously was closely similar to that for other CAPS, suggesting a similar balance of general practice and specialist products.

An advantage of the cross-sectional design used for this study lies in the ability to examine a range of products over a range of times from authorization that are typical of an established pharmacovigilance system. Thus, a product might enter the study at any point in its life-cycle. To show what this means in practice we present some additional characterization of study data. Figure 1 shows the numbers of substances in the study by calendar month over the time window for the study, while figure 2 shows the number of substances under study by month from first authorization. The area under these graphs shows the total information in the study to reflect 437 product years.

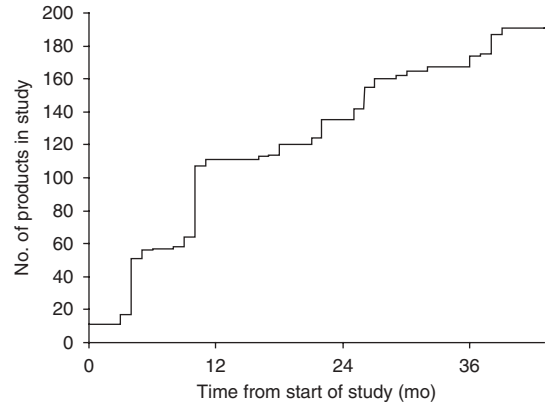

Fig. 1. No. of substances in the study by calendar month over the time window for the study.

Effectiveness of PRR Signal Detection

The 102 products for which type II variations were submitted during the study period contributed to this part of the study. These variations included 532 newly identified ADRs. Of these, 405 (76%) were on the list of IMEs.

The primary purpose of this study was to quantify the benefit that can be obtained by adding PRR signal detection to established pharmacovigilance methods. Thus, we are not concerned solely with whether PRR methods can detect ADRs but with whether they can detect ADRs earlier than the other methods.

The question of exactly which procedures can be considered to constitute 'established pharmacovigilance' is addressed only indirectly by this study. The fact is that the type II variations were the result of the full armamentarium of pharmacovigilance processes that was in place during and immediately before the study period, and it is this output, rather than a characterization of the processes that led to it, which is needed for the study. However, some insight into the established processes can be provided by the primary sources of information that have been identified as the precursors of the regulatory actions. This identification was carried out for the subset of 405 IMEs. A list of sources is given in table I.

Periodic Safety Update Reports (PSURs) provide summaries of current safety information for marketed medicines at pre-specified intervals.

By 'review' we mean a systematic evaluation either of the product or of some specific clinical area. 'Surveillance' encompasses any formal processes put in place either at the time of authorization or at an extension of indication because of a concern that insufficient evidence relating to the product was available. We define 'SRS cases' as those arising from review of spontaneous reporting systems other than EudraVigilance, but not from a PSUR or one of the other sources.

The principal results of the study with respect to detecting the 405 IMEs identified by established pharmacovigilance methods are shown in table II.

Using the sets of MedDRA synonyms, the ADR was picked up before the index date for standard pharmacovigilance in 217 cases. This represents 53.6% of all IMEs (95% CI 48.7, 58.4). The mean time gain in these cases was 2.45 years, and the range was from 0 to 8.34 years. The mean time gain when only exact MedDRA matches are employed is almost identical, but only 183 index dates were pre-empted by PRR-based detection (table III), representing 45.6% IMEs (95% CI 40.8, 50.5)

The distribution of time differences between detection by PRR methods and the index date for standard pharmacovigilance techniques can be presented graphically as a Kaplan-Meier curve. This is shown for the detection procedure using the MedDRA PT synonyms in figure 3. This graph uses the index date as reference point – set to zero – and hence the curve to the left of zero represents cases where the SDR preceded the index date. From the graph it is possible to see the proportion of ADRs for which any chosen time gain in detection was achieved.

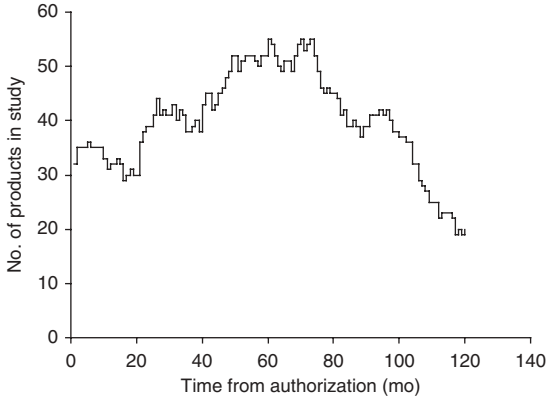

Fig. 2. No. of substances under study by month from first authorization.

Table I. Sources of data used to determine index date for established pharmacovigilance signals related to important medical events

Source	Frequency (%)
Periodic Safety Update Reports	259 (64.0)
Class effects	76 (18.8)
Surveillance	40 (9.9)
Clinical trial	15 (3.7)
Review	8 (2.0)
Other SRS[a]	4 (1.0)
US FDA	3 (0.7)

a Spontaneous reporting system (SRS) cases are defined as those arising from review of SRSs other than EudraVigilance, but not from a Periodic Safety Update Report or one of the other sources.

Results of Study to Assess Workload Associated with PRR Signal Detection

The risk or regulatory resource cost of adopting the PRR procedure is the effort involved in assessing the many other SDRs that will arise but prove to be unrelated to any pharmacological effect of the product. During the study period, 6356 new SDRs related to IMEs were observed in 139 of the 191 products in the study. These were independently evaluated by two assessors and were classified into the categories suggested by Hochberg et al.[9] For ease of reference, these categories are shown in table IV.

A cross classification of these categories for the two researchers is shown in figure 4.

A third assessor reviewed the set of discrepancies between the initial assessors when only one of them had classified an SDR as a potential signal of interest (R). The final classification identified 1561 of the 6356 SDRs as new and potentially important associations. These SDRs, which represent a significant cost to the system, are those that cannot be quickly classified, for example those not already listed on the SPC or attributable to confounding by indication. These are subject to a full assessment often involving inspection of narrative fields in the reports and a search for additional data. At the time of writing, these 1561 SDRs were undergoing

Table II. Detection of drug-event pairs using standard proportional reporting ratio (PRR) signal criteria defining a signal of disproportionate reporting (SDR) with synonymous Medical Dictionary for Regulatory Activities (MedDRA®) terms restricted to important medical events

Outcome measure	SDR from PRR within study		No SDR during study
	PRR first	standard pharmacovigilance first	
Safety issues [n (%)]	217 (53.6)	79 (19.5)	109 (26.9)
Mean time gain (y)	2.45	−1.07	
Maximum time gain (y)	8.34		

further review to identify any product safety concerns; the results of this procedure are not required for calculation of outcomes from this study.

As a measure of the incremental value of an additional signal detection method we have calculated the number of SDRs requiring full assessment per safety issue, which is first detected by the new method. In the case of the PRR method added to standard pharmacovigilance methods, we estimate this number to be 1561/217 = 7.2 (95% CI 6.5, 7.9)

This figure is an average over the entire study and hence approximately over a 10-year product life-span. To obtain an idea of how the return on statistical signal detection may vary over the life of the product it is instructive to look at the distribution over these 10 years of regulatory actions and SDRs. For the study data, the smoothed ratio of these quantities is shown graphically in figure 5. This graph indicates that statistical signal detection may be most efficient during the first 1 or 2 years of a product life cycle, but seems to reach a fairly stable figure of eight evaluations per signal pre-empted after that time.

The Effect of the Seriousness Filter

In performing this study we have used a list of IMEs to pre-filter our results. This list did not arise from the study but was already in the late stages of development in response to a perceived need for a filtering mechanism to focus attention on serious events. Such a list is an inescapable requirement for some commonly used statistical signal detection routines but is optional for our PRR-based method; thus, we were interested in evaluating its effect on our results.

Currently, 84% of reports to EudraVigilance include at least one IME; therefore, it might be thought that the use of this list would have little effect on the study. If it had not been used, events judged to be clinically unimportant would be quickly eliminated from the SDR list at the preliminary inspection phase and it would also seem unlikely that such events would have been added to the SPCs for each of the products.

In fact neither assumption is correct. This list reduced the number of MedDRA PTs considered from about 18 200 to 9078; therefore, whilst it is probably true that the extra SDRs would have been eliminated by inspection, the task would not have been insignificant because of the substantial numbers involved.

With respect to the second assumption, 127 of the 532 adverse events used in this study were not on the list and hence were classified as undetectable by the PRR methods. These involved a total of 79 different PTs. These were examined to

Table III. Detection of drug-event pairs using standard proportional reporting ratio (PRR) signal criteria defining a signal of disproportionate reporting (SDR) with exact Medical Dictionary for Regulatory Activities (MedDRA®) match only restricted to important medical events[a]

Outcome measure	SDR from PRR within study		No SDR during study
	PRR first	standard pharmacovigilance first	
Safety issues [n (%)]	183 (45.6)	84 (20.9)	134 (33.4)
Mean time gain (y)	2.43	−1.34	
Maximum time gain (y)	8.34		

a Note that four adverse events did not have exact MedDRA matches and could not be analysed.

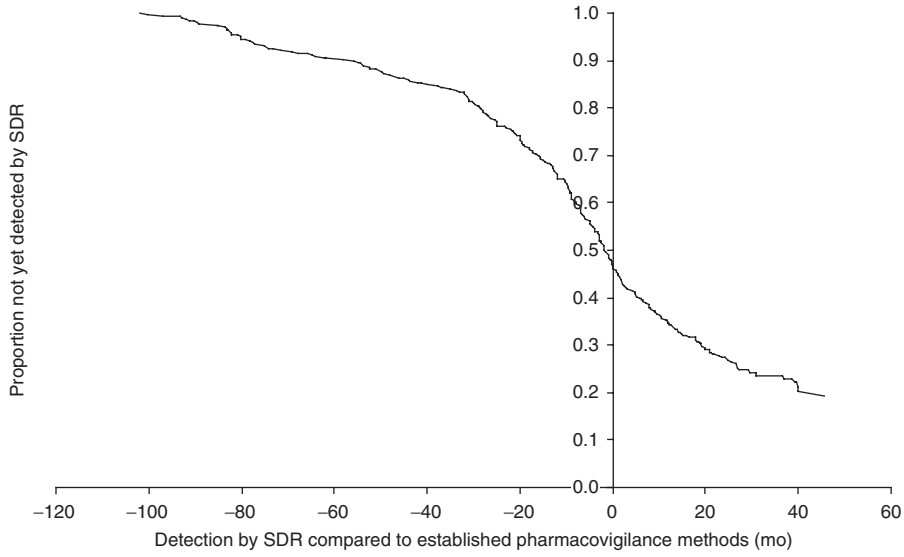

Fig. 3. Kaplan-Meier curve for the detection procedure using the Medical Dictionary of Regulatory Activities (MedDRA®) preferred term synonyms. **SDR** = signal of disproportionate reporting.

ensure that they did not include important events erroneously excluded from the IME list, but no such events were found. The reason that ADRs that do not require urgent action may be added to an SPC is usually that they are added in conjunction with a clinically related IME. Hence, failure to detect such reactions independently will make no difference to the eventual SPC.

Our conclusions from this are that the IME list could fulfill a role in reducing the overall workload of signal detection without missing important clinical signals, but it is obvious that a common IME list must be used when comparing different signal detection systems.

Classification of Drug-Event Pairs Included in Regulatory Actions not Predicted by PRR Methods

In order to understand more fully the way in which statistical signal detection fits into the wider pharmacovigilance system we looked in detail at the source of regulatory actions for which either no statistical signal arose or for which the signal arose only after the safety issue had arisen from another source.

We excluded from this review those 127 events that do not qualify as IMEs. Of the remaining 405 adverse events, 217 (53.6%) were predicted by an SDR arising prior to the signal from standard pharmacovigilance measures. This leaves 188 events that either raised an SDR after the standard measures (79 [19.5%]) or did not raise an SDR during the study period (109 [26.9%]).

Examination of the 188 events revealed that 30 were clinically associated with another event that had been predicted by an SDR and added to the SPC. In other words, these might be considered as part of a syndrome detected using the PRR. A further 42 were added on the basis that they were class effects. It is possible that these may have earlier resulted in raised SDRs for some other

Table IV. Classification of signals of disproportionate reporting[a]

Label	Description
L	Labelled in the summary of product characteristics
I	Indication for the medicinal product
D	Demographic features of the target population
G	General term – cannot be interpreted
R	Signal of interest

a Modified classification of Hochberg et al.[9]

		Assessor 1						
		D	G	I	L	R	Total	Agreement (%)[1]
Assessor 2	D	509	233	122	778	960	2602	**19.6**
	G	15	99	7	87	27	235	**42.1**
	I	7	0	49	4	1	61	**80.3**
	L	42	63	13	1832	226	2176	**84.2**
	R	39	175	17	319	732	1282	**57.1**
	Total	612	570	208	3020	1946	**6356**	
	Agreement (%)[1]	**83.2**	**17.4**	**23.6**	**60.7**	**37.6**		

Fig. 4. Comparison of classifications of signals of disproportionate reporting by two assessors. **1** Total agreement: Kappa = 0.33.

member of the product class, but this would take considerable additional work to verify. Another 32 events were found to have arisen from an active surveillance measure put in place at the time of authorization, and hence the date of signal detection for standard pharmacovigilance was the authorization date. These results are shown in figure 6.

Thus, 84 (20.7% of 405) events not predicted by an SDR remained, of which five arose directly from randomized controlled trials and two from reviews that concentrated on syndromes rather than PTs, suggesting that *ad hoc* grouping of terms may sometimes pay dividends in signal detection. Of the 77 (19.0%) unexplained events, 46 had more than three reports in EudraVigilance and hence could technically have resulted in a raised SDR.

Discussion

In a wide-ranging review of data-mining methods in pharmacovigilance, Almenoff et al.[10] discussed the difficulty of adapting outcome measures based on analogy with diagnostic studies to signal detection. In designing this study, we chose to abandon this analogy in favour of direct validation of the method. In other words, we chose to assess directly the extent to which our statistical methods could achieve their goal of finding safety problems earlier. In doing so we used a large group of drugs selected without reference to drug safety history and included all ADRs for these products within our study time window. The other side of the benefit-risk evaluation was represented by all the SDRs for these

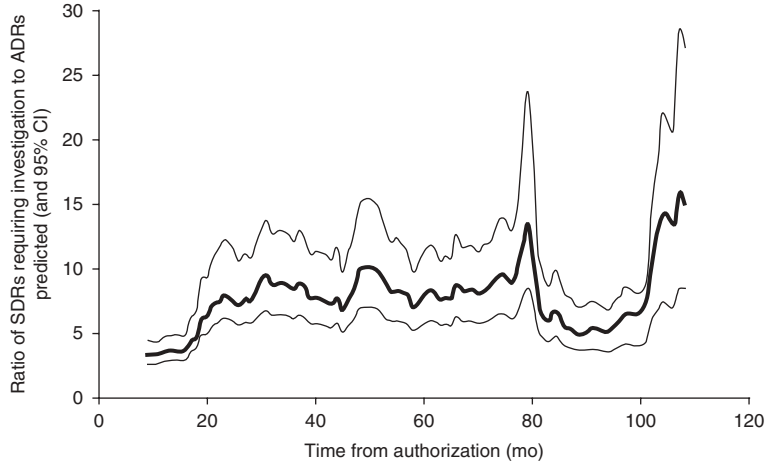

Fig. 5. No. of signals of disproportionate reporting (SDRs) investigated per adverse drug reaction (ADR) identified earlier – variation with time from authorization.

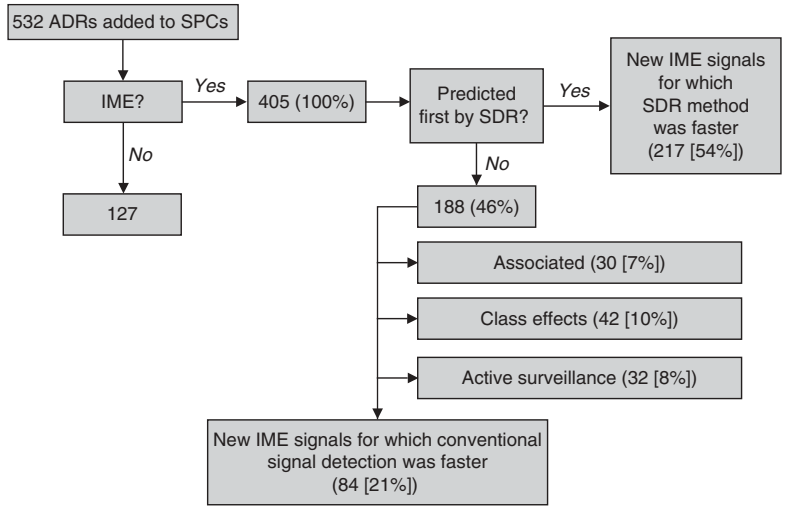

Fig. 6. Contributions of signal of disproportionate reporting (SDR) methods, established signal detection and other pharmacovigilance processes to detection of adverse drug reactions (ADRs). **IME** = important medical event; **SPC** = summary of product characteristics.

products arising within the same time window. In this way we have attempted to make a pragmatic assessment of the impact of statistical signal detection on a real-life pharmacovigilance system. Almenoff et al.[10] also noted that not all safety problems are picked up by disproportionality measures, and we believe that our review of drug-event pairs without early SDRs in our study may be the first systematic investigation of this observation.

Using US FDA Adverse Event Reporting System data and a different disproportionality-based signal detection method, Hochberg et al.[9] achieved results similar to those presented in this study. They examined 21 products authorized in 2001 and found that SDRs preceded regulatory action in 39/72 (54%) cases. The number of potential signals requiring evaluation was 244, giving a ratio of 6.3 per action pre-empted. Despite the numerical similarity of these results to our own, they are not directly comparable to ours for a number of reasons. First, we have put the statistical signal detection routines to a harder test. While Hochberg et al.[9] used the date of FDA regulatory activity as the index date, we have been able to trace the first dates at which the EMA became aware of the potential problem. This was often many months before the regulatory decision. Second, Hochberg et al.[9] only examined the first 3 years after authorization. This meant that the potentially large time gains in older products could not be investigated and that generalizations to routine pharmacovigilance (which continues over the lifetime of the product) could not be so robust. Our work on older products has shown that some signals that are detectable by disproportionality statistics can go many years undetected by other methods. Lastly, the back-population of EudraVigilance has meant that we have had a dataset that has not already been subjected in its entirety to statistical screening. Thus, the pharmacovigilance methods resulting in EMA regulatory actions during the study are likely to be qualitatively different from SDR methods; therefore, our study should give a realistic assessment of the added value of the PRR method. We have also extended this work with a much larger set of products and an investigation of the reasons why SDR methods sometimes fail to find an ADR earlier than other methods.

An interesting finding in assessing the many positive SDRs was the low level of inter-observer agreement concerning whether they required in-depth investigation. This raises the question of whether an SDR related to a true ADR will

always be selected for further evaluation. Our belief is that it is very likely to be selected provided, as in our study, the threshold of suspicion is set quite low. This is reflected in the fact that while 732 SDRs were rated as 'R' by both initial assessors, the final number selected for further evaluation was 1561. Thus, the adjudicator erred on the side of caution, in particular in the most difficult decisions regarding discrimination between events associated with the target population and the product. Since it is difficult to find assessors who are simultaneously competent to make this classification but unaware of the known ADRs for the products, conclusive empirical investigation of this point is not possible in a retrospective study. A strength of our evaluation of the SDRs is that it was designed to produce a process analogous in rigor to that used in routine pharmacovigilance (since that is what we were testing) and we believe that this criterion was fulfilled. Hence, the primary measure of evaluations per successful early detection will reflect what is achievable in everyday signal detection. We have not tested whether better agreement can be obtained with more detailed definitions and training.

Despite the pragmatic approach we have used, there are many issues we have not resolved and which remain the object of further work. We have examined only one of a number of possible disproportionality statistics. We have not adjusted for potential confounders or modelled the known ADRs in the dataset. We have not examined the influence of data coding errors, the completeness of backlog reporting or the assumption that the decision to code a drug as 'suspect' or 'concomitant' can be validly made on the basis of a single case report. Nonetheless, even with these several uncertainties, it appears reasonable to infer from our results that a simple statistical method such as the PRR can add significant benefits to pharmacovigilance processes, leading to more effective public health protection.

When designing this study we considered both a retrospective and a prospective design. Partly for reasons of timing, the retrospective study was preferred; however, a system of screening such as this should be the subject of an ongoing programme of quality assurance (QA). If designed with appropriate controls, over time, the data from such a QA programme could be compiled into a prospective evaluation of the system. This would also allow investigation of the accuracy of clinical assessment of SDRs as discussed above.

The analysis of this study focuses on the question of what value is added to established pharmacovigilance methods by the use of a disproportionality method. The question of whether PRR screening is a useful replacement for the established procedures is not addressed. It would be technically possible to address this latter question using the same data but the assumption would have to be made that the identification of an ADR and addition to the SPC for a product did not substantively increase or decrease the rate of reporting to EudraVigilance, and hence alter the chance that an SDR would arise. This seems quite unlikely and hence the analysis would lack credibility. In any case, the results of our current analysis clearly do not support a case to reduce other forms of effective pharmacovigilance activity.

The dataset we have compiled in the course of this study is suitable for evaluating other forms of signal detection analysis. Thus, the study dataset will be a useful tool for evaluation of any proposed alternative statistical screening techniques or changes of procedure, and will provide a direct comparison with the PRR. Work on the data is currently planned to further characterize the factors that predispose an ADR to detection using SDR methods, and to examine the impact of exclusion of drugs classified as concomitant and problems with data quality. We will also address the optimization of MedDRA coding levels and definitions of SDRs to maximize the net benefits from these screening techniques.

Conclusions

The screening technique based on the PRR can be used as an addition to established pharmacovigilance processes and has the potential to provide significant early warning in approximately 54% of cases where a clinically important ADR is found. This will allow earlier

investigation and the possibility of early regulatory action to reduce risks to public health. For each case in which an advantage is obtained in terms of earlier detection, we estimate that about seven investigations would be initiated, six of which would be evaluated as false positives. In cases where early detection is possible, significant delays can be avoided; on average, we have estimated these to be around 2.45 years.

A very important result of this study is that, whilst it demonstrates the strong supportive role that can be played in pharmacovigilance by statistical signal detection, it also shows clearly the importance of the established pharmacovigilance systems. Many signals are detected, first from active surveillance, reviews, clinical trials and PSURs, and instigation of additional measures should not be seen as a reason to reduce other pharmacovigilance activities.

Acknowledgements

This study was requested by the EudraVigilance Steering Committee and we would like to thank N. Wathion, P. Arlett, S. Brosch, H-G. Eichler and X. Kurz for their constructive comments.

The views expressed in this article are the personal views of the authors 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. The EMA does not profit financially from EudraVigilance or associated signal detection activities. None of the authors have any conflicts of interest to declare.

References

1. The rules governing medicinal products in the European Union. Notice to applicants and regulatory guidelines for medicinal products for human use. Vol. 2 [online]. Available from URL: http://ec.europa.eu/enterprise/sectors/pharmaceuticals/documents/eudralex/vol-2/index_en.htm [Accessed 2010 Apr 13]
2. The rules governing medicinal products in the European Union. Guidelines on pharmacovigilance for medicinal products in human use. Vol. 9A [online]. Available from URL: http://ec.europa.eu/enterprise/pharmaceuticals/eudralex/vol-9/pdf/vol9a_09-2008.pdf [Accessed 2010 Apr 13]
3. VanPuijenbroek E, Bate A, Leufkens HGM, et al. A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions. Pharmacoepidemiol Drug Saf 2002; 11: 3-10
4. European Medicines Agency. EudraVigilance: pharmacovigilance in the European economic area [online]. Available from URL: http://eudravigilance.emea.europa.eu/highres.htm [Accessed 2010 Apr 13]
5. 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 [online]. Available from URL: http://ec.europa.eu/enterprise/pharmaceuticals/eudralex/vol-1/reg_1993_2309/reg_1993_2309_en.pdf [Accessed 2010 Apr 13]
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. European Medicines Agency, EudraVigilance Expert Working Group. Guideline on the use of statistical signal detection methods in the EudraVigilance data analysis system (EMEA/106464/2006 rev.1) [online]. Available from URL: http://www.emea.europa.eu/pdfs/human/phvwp/10646406enfin.pdf [Accessed 2010 Apr 13]
8. MedDRA MSSO [online]. Available from URL: http://www.meddramsso.com [Accessed 2010 Apr 13]
9. Hochberg AM, Reisinger SJ, Pearson RK, et al. Using data mining to predict safety actions from FDA Adverse Event Reporting System data. Drug Inf J 2007; 41: 633-43
10. Almenoff J, Tonning JM, Gould AL, et al. Perspectives on the use of data mining in pharmacovigilance. Drug Saf 2005; 28 (11): 981-1007

Correspondence: *Jim Slattery*, European Medicines Agency, 7 Westferry Circus, Canary Wharf, London E14 4HB, UK.
E-mail: Jim.Slattery@ema.europa.eu