ORIGINAL RESEARCH ARTICLE

Drug Safety 2007; 30 (8): 715-726
0114-5916/07/0008-0715/$44.95/0

Criteria Revision and Performance Comparison of Three Methods of Signal Detection Applied to the Spontaneous Reporting Database of a Pharmaceutical Manufacturer

Yasuyuki Matsushita,[1,2] *Yasufumi Kuroda*,[2] *Shinpei Niwa*,[2] *Satoshi Sonehara*,[2] *Chikuma Hamada*[1] and *Isao Yoshimura*[1]

1 Faculty of Engineering, Tokyo University of Science, Tokyo, Japan
2 Daiichi Sankyo Co. Ltd., Tokyo, Japan

Abstract

Background and objective: Several statistical methods exist for detecting signals of potential adverse drug reactions in spontaneous reporting databases. However, these signal-detection methods were developed using regulatory databases, which contain a far larger number of adverse event reports than the databases maintained by individual pharmaceutical manufacturers. Furthermore, the composition and quality of the spontaneous reporting databases differ between regulatory agencies and pharmaceutical companies. Thus, the signal-detection criteria proposed for regulatory use are considered to be inappropriate for pharmaceutical industry use without modification. The objective of this study was to revise the criteria for signal detection to make them suitable for use by pharmaceutical manufacturers.
Methods: A model comprising 40 drugs and 1000 adverse events was constructed based on a spontaneous reporting database provided by a pharmaceutical company and used in a simulation to investigate appropriate criteria for signal detection. In total, 1000 pseudo datasets were generated with this model, and three statistical methods (proportional reporting ratio [PRR], Bayesian Confidence Propagation Neural Network [BCPNN] and multi-item gamma Poisson shrinker [MGPS]) for signal detection were applied to each dataset. The sensitivity and specificity of each method were evaluated using these pseudo datasets. The optimum critical value for signal detection (i.e. the value that achieved the highest sensitivity with 95% specificity) was identified for each method. The optimum values were also examined with the adverse events classified into two categories according to frequency. The three original detection methods and their revised versions were applied to a real pharmaceutical company database to detect 173 known adverse reactions of four drugs.
Results: The 1000 pseudo datasets consisted of an average of 81 862 reports and 11 407 drug-event pairs, including 1192 adverse drug reactions. The sensitivities of PRR, BCPNN and MGPS methods were 49%, 45% and 26%, respectively, whereas their specificities were 95%, 99.6% and 99.99%, respectively; these sensitivities were unacceptably low for pharmaceutical manufacturers, whereas the specificities were acceptable. The highest sensitivity for each method,

obtained by changing critical values and maintaining specificity at 95%, was 44%, 62% and 62%, respectively. When adverse events were classified into two categories, sensitivities as high as 75% for regular events and 39% for rare events were achieved with the revised BCPNN method. The critical values of the information component minus two standard deviations (IC − 2SD) index of the revised BCPNN method were greater than −0.7 for regular events and greater than −0.6 for rare events. The revised BCPNN method yielded 51% sensitivity and 89% specificity for the real dataset.

Conclusion: A lower critical value may be needed when signal-detection methodology is applied to the spontaneous reporting databases of pharmaceutical manufacturers. For example, it is recommended that pharmaceutical manufacturers use the BCPNN method with IC − 2SD criteria of greater than −0.7 for regular events and greater than −0.6 for rare events.

Background

Although new pharmaceuticals are placed on the market after their safety and efficacy are confirmed in clinical studies, the safety information obtained in clinical studies is limited. Consequently, pharmaceutical companies and regulatory agencies have to continue to collect information and scrutinise the safety of a drug even after it is approved. An important role of the drug-monitoring process is the rapid detection of unknown adverse drug reactions and the formulation of effective countermeasures. The necessary data analysis includes the collection and assessment of information from spontaneous reports. Spontaneous reports are reports of adverse events (referred to in this paper simply as 'events') thought to be adverse drug reactions. The report is an unsolicited communication by a healthcare professional or consumer to a company, regulatory authority or other organisation.[1]

The WHO defines a signal as "reported information on a possible causal relationship between an adverse event and a drug, the relationship being unknown or incompletely documented previously."[2] Spontaneous reporting data are analysed in order to detect an adverse drug reaction signal for a drug-event combination based on the number of reports for the drug-event pair exceeding a prescribed number. An example of this approach had been adopted by the Uppsala Monitoring Centre of the WHO (WHO-UMC).[3]

However, a disadvantage of this type of analysis is that it detects too many signals for drug-event combinations that are not real adverse drug reactions.[3] Consequently, new methods of analysing spontaneous report data that improve on this issue have been developed and are being used by regulatory agencies in Europe and the US, the WHO-UMC and pharmaceutical companies. The analysis methods examined in this study are representative of these widely used signal-detection algorithms.[4-6]

Each of the signal-detection methods has its distinguishing features, and no single method is suitable for all circumstances. Although there have been various investigations examining the characteristics of these methods and the appropriate criteria for each method, no clear guidelines have been established for determining which method should be used according to each circumstance.[7-12]

Although both pharmaceutical companies and regulatory agencies require information on adverse events in the same manner, their circumstances and objectives are different. The spontaneous reporting databases that are used by pharmaceutical companies generally consist of fewer drugs and have fewer reported events than the spontaneous reporting databases used by regulatory agencies. Furthermore, pharmaceutical company databases tend to compile data from related drugs into an 'all other drugs' category (figure 1), which may mask significant drug-event relationships because of the high frequency of events associated with other drugs.[12]

Pharmaceutical companies have various means, such as pharmacological examination or scrutiny of clinical data, for examining whether an event is an adverse drug reaction or not. Therefore, the balance between sensitivity and specificity requirements

	Specific event E_j	All other events	Total
Specific drug D_i	n_{ij}	$n_{i+}-n_{ij}$	n_{i+}
All other drugs	$n_{+j}-n_{ij}$	$n_{++}-n_{i+}-n_{+j}+n_{ij}$	$n_{++}-n_{i+}$
Total	n_{+j}	$n_{++}-n_{+j}$	n_{++}

Fig. 1. Notation in a 2 × 2 contingency table for a specific drug-event pair. See appendix for definitions.

may differ between pharmaceutical companies and regulatory agencies. For example, the signal-detection method used by a pharmaceutical company may be required to maintain specificity at an acceptable level (e.g. ≥95%) while providing the greatest possible sensitivity.

Because of the sensitivity problems associated with the original signal-detection algorithms, these methods are considered to be inappropriate for use by pharmaceutical companies without suitable modification. An adjustment to the critical value may be required to make them suitable for the characteristics of the spontaneous reporting databases used by pharmaceutical manufacturers and enable them to provide the performance required by these companies.

This study investigated three different signal-detection algorithms and assessed which method may be the most useful to pharmaceutical companies given the unique characteristics of their databases and the focus of their pharmacovigilance efforts.

Methods

Study Strategy

A model and conditions were established, based on the realistic spontaneous reporting database of a pharmaceutical company with which several of the authors are affiliated. The performance of three typical signal-detection methods was evaluated by a Monte Carlo simulation. The evaluation showed that the performance of the existing three methods was unsatisfactory for the purposes of a pharmaceutical company.

The critical value established for each method was then modified and a receiver operating characteristic (ROC) curve was constructed. The characteristics of the methods were then compared. In the comparison, events were divided into two classes according to the event frequency (i.e. regular or rare), and it was evaluated whether the performance of the method improved when the critical value was changed. The critical value that provided the maximum sensitivity with a required 95% specificity, which was considered a realistic requirement, was then established, and the sensitivity obtained with this critical value was determined for each method.

Finally, each method was applied to the company's database to evaluate the performance of the method for four drugs whose overall adverse reaction profile was already well known, and the reproducibility of the simulation results was examined. Numerical calculations were performed using SAS version 8.2 (SAS Institute, Cary, NC, USA) software.

Signal-Detection Methods Compared in the Study

The signal-detection methods examined in this study were the proportional reporting ratio (PRR),[4] Bayesian Confidence Propagation Neural Network (BCPNN),[6] and multi-item gamma Poisson shrinker (MGPS).[5,13,14] With all of these methods, an index value is calculated for each drug-event pair, and if the index value exceeds a certain critical value, a signal is considered to be detected.

The indices for the three methods examined in this study were the PRR, information component minus two standard deviations (IC − 2SD) and the lower 5th percentile of the posterior distribution (EB05), respectively. The reporting frequency for the drug-event pairs (D_i-E_j; with D_i indicating the drug and E_j the event) is expressed using the notation shown in figure 1 (n_{ij}, n_{i+}, n_{+j}, n_{++}). The definitions of these indices are shown in the appendix. The standard critical values adopted by organisations such as regulatory agencies and the WHO-UMC are shown in table I. The original PRR method also uses an independent chi-squared (χ^2) statistic and n_{ij} as index, as indicated in table I, in addition to the index referred to as PRR. However, in this study, the critical value of χ^2 was fixed at 4, and only the PRR critical value was changed. Moreover, with the original MGPS method, the data are

Table I. Standard signal-detection criteria[a]

Method	Index	Criterion for signalising
PRR	PRR	PRR >2 and χ^2 >4
BCPNN	IC − 2SD	IC − 2SD >0
MGPS	EB05	EB05 >2

a Refer to appendix for definitions of indices.

BCPNN = Bayesian Confidence Propagation Neural Network; **EB05** = lower 5th percentile of the posterior distribution; **IC − 2SD** = information component minus two standard deviations; **MGPS** = multi-item gamma Poisson shrinker; **PRR** = proportional reporting ratio; χ^2 = chi-squared.

stratified according to sex, age and reporting year and the method also includes higher order combinations of drugs and events, but in this study the strata were not used and higher order combinations were not the focus. The rationale for these approaches is provided in the discussion.

As the definitions in the appendix indicate, for drugs or events with no reports, the index cannot be calculated with any of the methods. In this study, no signal was considered to have been detected for such drugs and events.

With the PRR method, if the number of reports of the event of interest is zero for drugs other than the drug of interest (i.e. $n_{+j} - n_{ij} = 0$), the PRR index cannot be calculated. For such drug-event pairs, no signal was considered to have been detected with the PRR method.

Simulation Model

In the simulation, a spontaneous reporting model comprising 40 drugs and 1000 events was assumed. Random numbers were used to generate 1000 pseudo datasets, in which the number of reports for pairs (n_{ij}) of D_i drugs ($i = 1, 2, \ldots, 40$) and E_j events ($j = 1, 2, \ldots, 1000$) was assumed to be the actual value of a random variable for which D_i and E_j each independently followed a Poisson distribution $\mathrm{Po}(t_i\pi_{ij})$.

Here, t_i is the number of prescriptions of D_i. The distribution assumed for t_i in the simulation is as follows: for 2 drugs, t_i is 500 000; for 8 drugs, t_i is 250 000; for 10 drugs, t_i is 100 000; and for 20 drugs, t_i is 50 000. π_{ij} is the intensity of event E_j when drug D_i is prescribed, and for spontaneous intensity π_{oj}, $\pi_{ij} = k_{ij}\,\pi_{oj}$. Spontaneous intensity (π_{oj}) is the event incidence per prescription, and the distribution of π_{oj} that was assumed in the simulation is shown in table II. If the event was not an adverse drug reaction, the value of k_{ij} was 1; setting k_{ij} to a value >1 indicated that E_j was an adverse drug reaction to D_i.

In table II, there are 100 events for which π_{oj} is 1/1000 or greater. These events are referred to below as regular events, and the remaining 900 events are referred to as rare events. Events such as diarrhoea and headache were assumed as regular events, and events such as Stevens-Johnson syndrome were assumed as rare events.

For each drug D_i, 20 regular events and 20 rare events were randomly selected, with 1 value from the set {2, 3, 5, 10} selected with equal probability and used as the value of k_{ij}. Thus, 40 adverse drug reactions were determined for each drug. This number of adverse drug reactions was determined based on the numbers of adverse drug reactions indicated in many package inserts.

According to Lawson et al.,[15] the events reported through spontaneous reporting represent between 1/50 and 1/10 of the events caused by drug treatment. The simulation took this into account by decreasing the intensity (π_{ij}) to 1/50 when generating the hypothetical data (n_{ij}).

Indices for Performance Evaluation

The model included 40 adverse drug reactions for each drug. For all of the drugs combined, there were 1600 drug-adverse reaction pairs (40×40). When each method was applied to the pseudo datasets generated in the simulation, the drug-event pairs for which a signal was detected were determined. Of the 1600 drug-adverse reaction pairs, the proportion for which a signal was detected indicated the sensitivity, in its true sense. However, for drugs with a small number of prescriptions and events with a low spontaneous intensity, it was unlikely that problematic pairs would occur. Consequently, of the 1600 drug-adverse reaction pairs examined, only those for which there was at least one report were included in the denominator in the sensitivity calculation for the performance evaluation of each method. Under this definition, the mean sensitivity for the 1000 pseudo datasets was considered to be the sensitivity of each method.

The number of pairs consisting of a drug and an event that was not an adverse drug reaction was 38 400 (40 × 960). Of these drug-event pairs, those for which there was at least one report were included in the denominator of the specificity calculation, and the proportion of these for which a signal was not detected was considered the specificity. The mean specificity for the 1000 pseudo datasets was considered the specificity of each method.

The sensitivity and specificity with the 1000 events divided into regular events and rare events were also evaluated for each method. This approach was taken because the method with the best performance could have differed depending on the frequency of events.

Revision of Critical Values

There is a trade-off relationship between sensitivity and specificity. Changing the critical value causes sensitivity and specificity to change in tandem. As was mentioned previously, from the perspective of a pharmaceutical company, it is reasonable to adjust the critical value to maximise sensitivity under the condition that specificity is maintained at 95%.

Therefore, the critical value at 95% specificity was determined by changing the critical value in small increments and constructing a ROC curve, in which sensitivity was plotted along the vertical axis and '1 – specificity' along the horizontal axis, for all events combined and with the events divided into classes. It was thought that for a pharmaceutical company, of the methods for which the critical value was revised as previously described (i.e. revised PRR, BCPNN and MGPS methods), it would be valid to use the method that provided the highest sensitivity.

Comparison of Signal-Detection Methods Using a Real Dataset

Some of the authors are affiliated with a pharmaceutical company and were therefore able to use its spontaneous reporting databases. Individual case safety reports from spontaneous reports, medical literature and postmarketing studies were included in the database. The three methods examined and the revised versions of those methods were applied to a portion of that database (referred to as the 'real dataset') to determine whether the performance comparison results obtained in the simulation could be reproduced with real data.

The real dataset comprised a total of 19 559 reports, 34 959 reports aggregating drug-event pairs, 82 drugs, 1690 reported events and 5864 drug-event pairs included in reports. The number of reports for each drug-event pair was 1 for 58% of all pairs, 2 for 14%, 3–5 for 13% and ≥6 for 14%.

The 82 drugs in the real dataset included four drugs that had an adverse drug reaction profile that was well known and that had been on the market for ≥10 years (two therapeutic agents for chronic disease and two therapeutic agents for acute disease). The sensitivity of the methods was evaluated by determining whether signals were detected for the 173 drug-adverse reaction pairs indicated in the package inserts for these drugs. The adverse events were classified according to the Medical Dictionary for Regulatory Activities (MedDRA), but the terminology used in the real dataset corresponded to the terminology used in the package inserts.

Results

Summary of Generated Pseudo Data

For the 1000 pseudo datasets generated, the mean total number of reports (n_{++}) was 81 862, the mean number of drug-event pairs reported was 11 407, and the mean number of drug-adverse reaction pairs was 1192. When events were divided into two classes, regular events and rare events, the mean total number of reports was 69 330, the mean number of drug-event pairs reported was 3580, and the mean number of drug-adverse reaction pairs was 789 for regular events. For rare events, the mean total number of reports was 12 532, the mean number of drug-event pairs reported was 7828 and the mean number

Table II. Distribution of spontaneous intensities (π_{0j}) assumed in the simulation model comprising 40 drugs and 1000 events

π_{0j} (per 1000)	50	20	10	5	2	1	0.5	0.2	0.1	0.05
No. of events	2	3	5	15	25	50	100	100	200	500

Table III. Observed sensitivity and specificity of three signal-detection methods

Class of events	Method	Percentage sensitivity [mean (SD)]	Percentage specificity [mean (SD)]
All events	PRR	49 (1.3)	95 (0.4)
	BCPNN	45 (1.1)	99.6 (0.1)
	MGPS	26 (1.0)	99.99 (0.01)
Regular events	PRR	52 (1.4)	99.7 (0.1)
	BCPNN	57 (1.4)	99.8 (0.1)
	MGPS	33 (1.2)	99.999 (0.007)
Rare events	PRR	44 (2.3)	93 (0.5)
	BCPNN	21 (1.9)	99.5 (0.1)
	MGPS	12 (1.5)	99.98 (0.02)

BCPNN = Bayesian Confidence Propagation Neural Network; **MGPS** = multi-item gamma Poisson shrinker; **PRR** = proportional reporting ratio.

of drug-adverse reaction pairs was 403. Thus, the total number of reports was approximately 5-fold higher for regular events than for rare events, the number of drug-event pairs reported for regular events was approximately half the number for rare events, and the number of drug-adverse reaction pairs was approximately 2-fold higher for regular events than for rare events.

The number of reports for each drug-event pair was 1 for 53% of all pairs, 2 for 16%, 3–5 for 15% and ≥6 for 16%. This is validated in the discussion section.

Among the 1000 pseudo datasets, a PRR value could not be calculated for 0.14% of the drug-event pairs. The mean values of the hyperparameters (α_1, β_1, α_2, β_2, p) in the 1000 datasets for the MGPS were 2.44, 0.93, 8.97, 14.0 and 0.17, respectively.

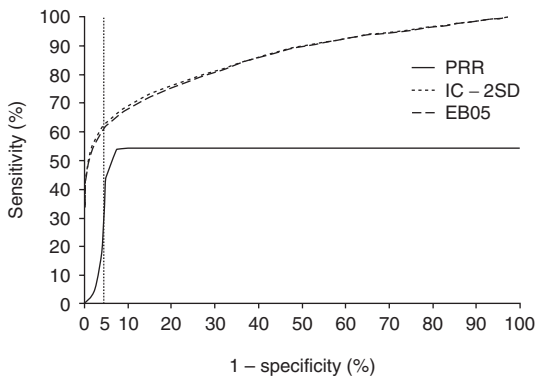

Fig. 2. Receiver operating characteristic curves for all adverse events. **EB05** = the lower 5th percentile of the posterior distribution; **IC – 2SD** = information component minus two standard deviations; **PRR** = proportional reporting ratio.

Sensitivity and Specificity of Each Method

The sensitivities and specificities of the three methods when applied to the 1000 pseudo datasets are shown in table III. Sensitivity for the 1000 events overall was highest with the PRR method (49%) and lowest with the MGPS method (26%). Specificity was 95% with the PRR method and approximately 100% with the other methods.

When the events were divided into two classes, sensitivity was higher for regular events than for rare events with all three methods. Specificity was approximately 100% for all methods except the PRR method to detect rare events. Sensitivity was highest at 57%, when the BCPNN method was used to detect regular events, and lowest at 44%, when the PRR method was used to detect rare events.

Evaluation by Receiver Operating Characteristic Curve

The ROC curve for all 1000 events is shown in figure 2, the curve for regular events is shown in figure 3, and the curve for rare events is shown in figure 4. Because the shape of the ROC curves for regular events and rare events differ, it was concluded that the critical values should be revised independently for each event class. As figures 2, 3 and 4 all show, the ROC curve for the PRR method differed from the curves for the other methods in that sensitivity with the PRR method had a tendency to pla-

teau at a certain value. The reason for this is discussed in the discussion section.

Effect of Classification of Events

The critical values and sensitivities of the methods when specificity was 95% are shown in table IV. With the PRR method, the critical value varied widely between classes. In the evaluation by event class, sensitivity was highest for the BCPNN method. Sensitivity with this method was 75% for regular events and 39% for rare events.

Overall Sensitivities Achieved with the Revised Methods

Based on these results, the critical value was adjusted for each of the two event classes so that an overall specificity of 95% was maintained, and the results shown in table V were obtained. The highest sensitivity, 64%, was observed with the revised BCPNN method, and it was concluded that this is the reasonable signal-detection method for a pharmaceutical company.

Application of Six Methods to a Real Dataset

The number of drug-event pairs for the four drugs for which there were reports in the real dataset and that were examined in this study was 1932. Of the 173 drug-adverse reaction pairs indicated in the package inserts, there were eight pairs with no reports in the real dataset. The sensitivities and specificities of the three methods examined in this study and their revised versions are shown in table VI. With the revised BCPNN method, sensitivity was 51% and specificity 89%. Although the performance of this method was inferior to that shown by the simulation, it was relatively superior when the six methods were compared.

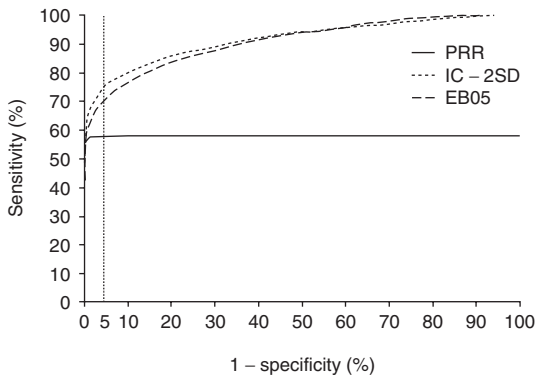

Fig. 3. Receiver operating characteristic curves for regular adverse events. **EB05** = the lower 5th percentile of the posterior distribution; **IC – 2SD** = information component minus two standard deviations; **PRR** = proportional reporting ratio.

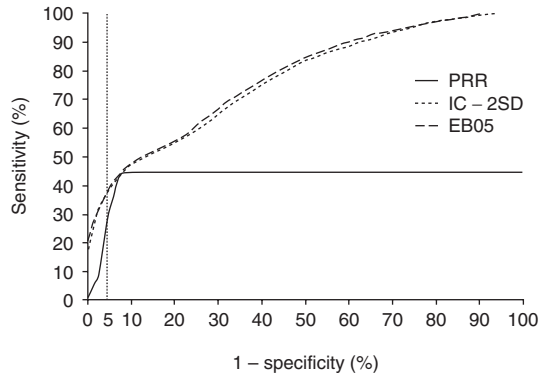

Fig. 4. Receiver operating characteristic curves for rare adverse events. **EB05** = the lower 5th percentile of the posterior distribution; **IC – 2SD** = information component minus two standard deviations; **PRR** = proportional reporting ratio.

The relative relationships between the performance of the six methods were generally agreed upon with respect to the simulation results. However, the sensitivities and specificities were lower with the real dataset. This was possibly attributable to the fact that the total number of reports in the real dataset was approximately a quarter of the number used in the simulation.

Discussion

Generalisability of the Simulation Model

The spontaneous reporting databases that were used to develop many of the signal-detection methods are huge. The spontaneous reporting database of the WHO-UMC, which gathers data from regulatory agencies in various countries, contains >3.4 million reports,[12] and the spontaneous reporting database of the US FDA contains >2.6 million reports.[12] Among large pharmaceutical companies, Schering AG has a spontaneous reporting database that con-

Table IV. Critical value and sensitivity with 95% specificity

Class of events	Method	Critical value [mean (SD)]	Percentage sensitivity [mean (SD)]
All events	PRR	2.3 (1.67)	44 (10.1)
	BCPNN	−0.61 (0.02)	62 (1.3)
	MGPS	0.55 (0.01)	62 (1.4)
Regular events	PRR	0.53 (0.02)	59 (1.5)
	BCPNN	−0.64 (0.04)	75 (1.6)
	MGPS	0.59 (0.02)	71 (1.8)
Rare events	PRR	8.2 (1.36)	30 (3.8)
	BCPNN	−0.60 (0.02)	39 (2.4)
	MGPS	0.53 (0.01)	39 (2.4)

BCPNN = Bayesian Confidence Propagation Neural Network; **MGPS** = multi-item gamma Poisson shrinker; **PRR** = proportional reporting ratio.

tains >230 000 reports, but most pharmaceutical companies do not have spontaneous reporting databases of that scale.[10]

The mean number of reports in the pseudo datasets generated in this study was 81 632, which corresponds to the size of a spontaneous reporting database from a moderate-size pharmaceutical company. Therefore, the revised BCPNN method, which provided the highest sensitivity in the simulation performed in this study, appears to be a suitable method for analysing spontaneous reporting databases from moderate-size pharmaceutical companies.

Regarding the distribution of the number of reports for each drug-event pair, the spontaneous reporting database of the regulatory agency in the UK has one report for approximately 60% of pairs,[16] the spontaneous reporting database of the regulatory agency in the Netherlands has one report for 68% of pairs and two reports for 14%,[7] and the spontaneous reporting database of the regulatory agency of Japan has one report for 65% and two reports for 16% of pairs.[8]

In the pseudo datasets generated in this study, there was one report for 53% of pairs and two reports for 16% of pairs, for a combined proportion of approximately 70%. These proportions are comparable to the real world examples, indicating that the pseudo datasets used in this study filled the general conditions of a spontaneous reporting database with respect to the distribution of the number of reports.

In actual situations, the reporting rate is dependent on many factors including the time since launch, pharmacovigilance-related regulatory activity, media attention and the indication for use of the drug.[17] Such conditions were not incorporated into the simulation model in this study, which limits the generality of the conclusion of the study.

Desirable Performance for Pharmaceutical Manufacturers

When the existing signal-detection methods were used, specificity was around 100% and sensitivity was <50% with all of the methods. When the criterion n_{ij} >2 was added, the PRR method still provided specificity of 99%. The high specificity and low

Table V. Critical value and overall sensitivity of revised methods with 95% specificity

Method	Criteria		Percentage sensitivity	Percentage specificity
	regular events	rare events	[mean (SD)]	[mean (SD)]
Revised PRR	PRR >0.5	PRR >8.2	49 (1.4)	95 (0.3)
Revised BCPNN	IC − 2SD greater than −0.7	IC − 2SD greater than −0.6	64 (1.3)	95 (0.3)
Revised MGPS	EB05 >0.59	EB05 >0.53	60 (1.5)	95 (0.6)

BCPNN = Bayesian Confidence Propagation Neural Network; **EB05** = the lower 5th percentile of the posterior distribution; **IC − 2SD** = information component minus two standard deviations; **MGPS** = multi-item gamma Poisson shrinker; **PRR** = proportional reporting ratio.

sensitivity of these existing signal-detection methods reflects the fact that they were developed to examine serious events as part of a prioritisation approach.[16] Consistent with this approach, is the triage logic introduced by the WHO-UMC[9] and the impact analysis introduced by the UK Medicines and Healthcare products Regulatory Agency (MHRA),[18] which give a priority ranking to the examination of detected signals.

For pharmaceutical companies, it is necessary from the perspective of risk management to detect adverse events suspected of being adverse drug reactions as early as possible and examine whether they truly are adverse drug reactions. In our opinion, even if it allows some increase of false-positives, the pharmaceutical company requires high-enough sensitivity to find missed true adverse drug reactions in daily manual case review. In this sense, pharmaceutical companies require a different method of signal detection from the prioritisation approach. This aspect was the main focus of this study.

Classification of Events by Frequency

The indices of the signal-detection methods examined in this study are determined essentially by evaluating the size of the increase in the frequency of an event in relation to its spontaneous intensity, expressed as a ratio. However, if the data follow a Poisson distribution and the spontaneous intensity is low, the absolute increase in number of occurrences has a greater effect on power than the ratio. Therefore, for an event with a high spontaneous intensity, it is appropriate to use an existing method that detects a signal by determining whether there is a multi-fold increase in incidence, but for an event with a low spontaneous intensity, it is more reasonable to examine the number of occurrences. From this standpoint, it is reasonable to change the critical value used for the signal-detection method according to the spontaneous intensity of the event.

Unrevised Criteria

The original PRR method includes the criteria n_{ij} >2 and χ^2 >4. However, in this study, the critical value of the χ^2 statistics was fixed at 4 and criterion n_{ij} >2 was deleted. This approach was used because including the n_{ij} >2 criterion resulted in too low sensitivity when specificity was 95%. This was attributable to the fact that, in an spontaneous reporting database with a small total number of reports, an increase in the number of events with one or two reports has a substantial effect and, therefore, the n_{ij} >2 criterion has too large an impact on sensitivity and specificity.

Table VI. Sensitivity and specificity obtained by six methods for a real dataset

Method	Sensitivity (%)	Specificity (%)
PRR	29	90
BCPNN	32	95
MGPS	12	99
Revised PRR	28	93
Revised BCPNN	51	89
Revised MGPS	56	82

BCPNN = Bayesian Confidence Propagation Neural Network; **MGPS** = multi-item gamma Poisson shrinker; **PRR** = proportional reporting ratio.

In some investigations, the χ^2 statistic has been adjusted at the same time. However, in cases where overall specificity of 95% is obtained, a χ^2 criterion of >4 is known to be reasonable and, therefore, the criterion was held constant at 4 in this study. The plateau seen in the ROC curves shown in figures 2, 3 and 4 is an effect of the χ^2 statistic, and changing the critical value for this did not achieve an upward expansion of the ROC curve as a whole. We also investigated another criterion of the PRR method; the 95% lower confidence limit of PRR (PRRCI) instead of χ^2 criterion. The overall sensitivity of revised PRRCI with 95% specificity (PRRCI >0.7 for regular events and PRRCI >2.4 for rare events) was 60%, which was lower than that of the revised BCPNN.

In the MGPS method, the FDA uses an estimated value for each stratum of variables such as sex and age to evaluate the expected value E_{ij}. Applying this approach to the moderate spontaneous reporting database of a pharmaceutical company would result in estimating hyperparameter due to the small total number of reports. An unstratified estimate of E_{ij} was used for the MGPS method in this study in order to avoid this instability. We did not consider higher order combinations for the same reason.

In the BCPNN method, we also investigated other indices; 80% or 90% lower confidence limits, and

the sensitivities with 95% specificity were 64% and 63%, respectively, which were similar as those of IC − 2SD. Therefore, we did not change the index of the BCPNN method.

Considering that the cut of the IC negative seems counterintuitive, the case with an additional criterion IC >0 was examined, the achieved sensitivity and specificity in real dataset being 49% and 89%, respectively, which were still superior to other methods.

Issues for Future Studies

In this study, events were divided into regular events and rare events based on the assumed spontaneous intensity. A practical problem that complicates this type of classification is the fact that the spontaneous intensity is not generally known. Although it is realistic to consider common events such as diarrhoea and headache and events occurring frequently in clinical trials to be regular events, the problem of how to divide events into specific classes remains a subject for future investigation.

In the model used for the simulation performed in this study, the number of adverse drug reactions per drug was established at 40, and their intensity was fixed at 2- to 10-fold higher than the spontaneous intensity. This differs from actuality in a number of ways. Whether the same performance in this study can also be obtained in such actual cases is a subject for future research.

Other subjects for future examination include an investigation of whether the extent of under-reporting depends on the setting (general practice or hospital setting), type of adverse drug reaction and type of drug;[19] an investigation of the effect of the duration of the marketing of a drug, as represented by the Weber effect;[20] an investigation of the MGPS method with stratification by age and sex;[5] and an investigation that considers abrupt increases in the number of reports.

The adverse drug reactions indicated in the package inserts are not a true gold standard, and the goal of the detection is to identify unknown adverse drug reactions. However, it is difficult to identify unknown adverse drug reactions in the real dataset, so we substituted a gold standard with the labelled adverse drug reactions of well characterised drugs. The validity of our strategy should be checked through daily pharmacovigilance activities.

Conclusion

A simulation was performed using a model based on the spontaneous reporting database of a pharmaceutical company, and the performance of the PRR, BCPNN and MGPS methods was evaluated. When the existing criteria were used, sensitivity was excessively low with all of the methods, which might make them unsuitable for the requirements of pharmaceutical companies. The critical values of the signal-detection method may need to be lowered when applied to the spontaneous reporting databases of pharmaceutical manufacturers. When the critical values were changed so that specificity was 95% and different critical values were used for regular events and rare events, it was shown that use of a revised BCPNN method, in which IC − 2SD was greater than −0.7 for regular events and greater than −0.6 for rare events, seemed to be the best signal-detection method to meet the needs of a pharmaceutical company.

Acknowledgements

The authors are grateful to Daiichi Sankyo Co., Ltd., for allowing access to its database. We would like to thank the two anonymous reviewers whose comments greatly improved this article.

Drs Chikuma Hamada and Isao Yoshimura have no conflicts of interest directly relevant to the content of this study. Yasuyuki Matsushita, Yasufumi Kuroda, Shinpei Niwa and Satoshi Sonehara are all employees of Daiichi Sankyo Co. Ltd.

No sources of funding were used to assist in the preparation of this study.

Appendix

Definitions of the indices used for the three methods (figure A1), using notation in table II, are as follows:

n_{ij} = number of reports of a drug (D_i) − event (E_j) pair;

n_{i+} = number of reports of a drug (D_i);

n_{+j} = number of reports of an event (E_j);

n_{++} = total number of reports.

1. Proportional reporting ratio (PRR) method (equation 1):

$$PRR = \frac{n_{ij}/n_{i+}}{(n_{+j} - n_{ij})/(n_{++} - n_{i+})}$$

$$\chi^2 = \frac{n_{++}\left(\left|n_{ij}(n_{++} - n_{i+} - n_{+j} + n_{ij}) - (n_{i+} - n_{ij})(n_{+j} - n_{ij})\right| - n_{++}/2\right)^2}{n_{i+}n_{+j}(n_{++} - n_{i+})(n_{++} - n_{+j})}$$

(Eq. 1)

Although $n_{ij} > 2$, in addition to PRR >2 and $\chi^2 > 4$, was used as a criterion for signal detection for a drug (D_i)-event (E_j) pair in the original proposal,[4] n_{ij} was not used as a criterion in this study (see discussion).

2. Bayesian Confidence Propagation Neural Network (BCPNN) method.
IC − 2SD = E − 2SD, where E, SD and γ are defined as follows (equation 2):[6]

$$E = \log_2 \frac{(n_{ij} + 1)(n_{++} + 2)(n_{++} + 2)}{(n_{++} + \gamma)(n_{i+} + 1)(n_{+j} + 1)}$$

$$SD^2 = \frac{1}{(\log_e 2)^2}\left(\frac{n_{++} - n_{ij} + \gamma - 1}{(n_{++} + 1)(1 + n_{++} + \gamma)} + \frac{n_{++} - n_{i+} + 1}{(n_{i+} + 1)(1 + n_{++} + 2)} + \frac{n_{++} - n_{+j} + 1}{(n_{+j} + 1)(1 + n_{++} + 2)}\right)$$

and

$$\gamma = \frac{(n_{++} + 2)(n_{++} + 2)}{(n_{i+} + 1)(n_{+j} + 1)}$$

(Eq. 2)

These formulas are derived from the originally proposed formulas[6] incorporating the usual assumptions and parameter estimates.

3. Multi-item gamma Poisson shrinker (MGPS) method (equation 3):

$$p\,\frac{\beta_1^{\alpha_1}\lambda_{ij}^{\alpha_1 - 1}}{\Gamma(\alpha_1)}\,e^{-\beta_1\lambda_{ij}} + (1 - p)\frac{\beta_2^{\alpha_2}\lambda_{ij}^{\alpha_2 - 1}}{\Gamma(\alpha_2)}\,e^{-\beta_2\lambda_{ij}}$$

(Eq. 3)

Let n_{ij} be a realised value of a random variable in a Poisson distribution, with mean $E_{ij}\,\lambda_{ij}$, where λ_{ij} is the intensity of occurrence of the drug-event pair D_i-E_{ij}; E_{ij} was assumed to be $n_{i+}\,n_{+j}/n_{++}$ in this investigation. When the prior distribution of λ_{ij} is a mixed γ distribution given in the above formula and the hyperparameters (α_1, β_1, α_2, β_2, p) in this distribution are estimated by an empirical Bayes method based on a spontaneous reporting database, we can obtain the posterior distribution of λ_{ij} given n_{ij}. The lower 5th percentile in this posterior distribution (EB05) is the index in the MGPS method.[5]

Fig. A1. Definitions of the indices used for the three signal-detection methods.[4-6] **IC − 2SD** = information component minus two standard deviations.

References

1. ICH-E2D-Expert Working Group. Post-approval safety data management: definitions and standards for expedited reporting [online]. Available from URL: http://www.ich.org/LOB/media/MEDIA 631.pdf [Accessed 2006 Jul 28]
2. Edwards IR, Biriell C. Harmonisation in pharmacovigilance. Drug Saf 1994; 10 (2): 93-102
3. Lindquist M, Edwards IR, Bate A, et al. From association to alert: a revised approach to international signal analysis. Pharmacoepidemiol Drug Saf 1999; 8: S15-25
4. Evans SJW, Waller P, Davis S. Proportional reporting ratios: the uses of epidemiological methods of signal generation. Pharmacoepidemiol Drug Saf 1998; 7: S102
5. Szarfman A, Machado SG, O'Neill RT. Use of screening algorithms and computer systems to efficiently signal higher-than-expected combinations of drugs and events in the US FDA's spontaneous reports database. Drug Saf 2002; 25 (6): 381-92

6. Bate A, Lindquist M, Edwards IR, et al. A Bayesian neural network method for adverse drug reaction signal generation. Eur J Clin Pharmacol 1998; 54: 315-21
7. van Puijenbroek EP, 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
8. Kubota K, Koide D, Hirai T. Comparison of data mining methodologies using Japanese spontaneous reports. Pharmacoepidemiol Drug Saf 2004; 13: 387-94
9. Stahl M, Lindquist M, Edwards IR, et al. Introducing triage logic as a new strategy for the detection of signals in the WHO Drug Monitoring Database. Pharmacoipidemiol Drug Saf 2004; 13: 355-63
10. Follmann M, Michel A, Geyer C. Comparison of different methods for signal detection in the drug-safety database of a pharmaceutical company. In: the 20th International Conference of Pharmacoepidemiology (ICPE) & Therapeutic Risk Management; 2004 Aug 22-4; Bordeaux. Pharmacoepidemiol Drug Saf 2004; 13: S111
11. Roux E, Thiessard F, Fourrier-Reglat A, et al. Evaluation of Statistical Association Measures for the Automatic Signal Generation in Pharmacovigilance. IEEE Trans Inf Technol Biomed 2005; 9 (4): 518-27
12. Almenoff J, Tonning JM, Gould AL, et al. Perspectives on the use of data mining in pharmacovigilance. Drug Saf 2005; 28 (11): 981-1007
13. DuMouchel W. Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. Am Statist 1999; 53 (3): 177-202
14. DuMouchel W, Pregibon D. Empirical bayes screening for multi-item associations. Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2001 Aug 26-9; San Francisco, 67-76
15. Lawson DH, O'Connor PC, Jick H. Drug attributed alterations in potassium handling in congestive cardiac failure. Eur J Clin Pharmacol 1982; 23: 21-5
16. Evans SJW. Statistics: analysis and presentation of safety data. In: Talbot J, Waller P, editors. Stephens' detection of new adverse drug reactions. 5th ed. West Sussex: Wiley, 2004: 316-25
17. ICH-E2E-Expert Working Group. Pharmacovigilance planning [online]. Available from URL: http://www.ich.org/LOB/media/MEDIA 1195.pdf [Accessed 2007 Jan 11]
18. Waller P, Heeley E, Moseley J. Impact analysis of signal detected from spontaneous adverse drug reaction reporting data. Drug Saf 2005; 28 (10): 843-50
19. Hazell L, Shakir SAW. Under-reporting of adverse drug reactions: a systematic review. Drug Saf 2006; 29 (5): 385-96
20. Weber JCP. Epidemiology of adverse reactions to non-steroidal antiinflammatory drugs. Adv Inflam Res 1984; 6: 1-7

Correspondence: Mr *Yasuyuki Matsushita*, Clinical Data and Biostatistics Department, R&D Division, Daiichi Sankyo Co. Ltd., 1-2-58, Hiromachi, Shinagawa-ku, Tokyo, 140-8710, Japan.
E-mail: matsushita.yasuyuki.ks@daiichisankyo.co.jp