SHORT COMMUNICATION

Drug Safety 2002; 25 (6): 423-432
0114-5916/02/0006-0423/$25.00/0

Automated Signal Generation in Prescription-Event Monitoring

Emma Heeley,[1] *Lynda V. Wilton*[1] and *Saad A.W. Shakir*[1,2]

1 Drug Safety Research Unit, Bursledon Hall, Southampton, United Kingdom
2 London School of Hygiene and Tropical Medicine, Public Health and Policy, London, United Kingdom

Abstract

Signal generation is a method of highlighting potential safety issues in a drug that then need to be investigated further. Previously automated signal generation has mainly been applied to spontaneous reporting systems. The Drug Safety Research Unit (DSRU) performs observational postmarketing studies on selected newly marketed medicines in England using a method known as prescription-event monitoring (PEM). The DSRU has investigated automated procedures for the generation of signals using the event data from PEM studies.

Proportional reporting ratios (PRRs) and incidence rate ratios (IRRs) were studied as possible tools for signal generation in PEM data. The PEM database contains 78 completed studies of drugs prescribed in primary care from a variety of therapeutic classes. Retrospective studies were carried out to identify the implications of changing the comparator group of drugs, along with analysing the results at different levels in the DSRU's hierarchical dictionary and performing signal generation after 30 and 180 days of observation since starting the drug.

Automated signal generation is a useful hypothesis generating method that is likely to prove to be useful both in clinical trials and postmarketing studies. PRRs are simple to apply and do not require a denominator. IRRs take into account the time subjects were exposed to the drug prior to the event of interest, and offers a useful, and more in depth look into the data. However, with both methods it is important to perform signal generation at multiple levels in the dictionary and with careful selection of the comparator group.

Identification of previously unrecognised adverse drug reactions is an important part of postmarketing surveillance. Automated signal generation is making a significant contribution to the identification of drug safety issues, which then need to be validated.[1,2] The Drug Safety Research Unit (DSRU) is an independent medical charity established in Southampton, England in 1980 following the thalidomide and practolol disasters. The unit monitors selected newly licensed products in the primary care setting in England by performing prescription-event monitoring (PEM) studies, which is possible because of the National Health Service (NHS) in the UK.

PEM is a noninterventional method of postmarketing surveillance whereby a cohort of approximately 10 000 patients are observed for at least 6 months after starting the drug of interest. A more detailed explanation of PEM methodology may be found elsewhere.[3] Patients are identified from NHS prescriptions written by primary care physicians [general practitioners (GPs)] and dispensed by pharmacists who are then reimbursed by the Prescription Pricing Authority (PPA). For the selected drugs monitored, the PPA sends prescription data to the DSRU in confidence. We subsequently send out questionnaires ('green forms') to the prescribing GPs asking for demographic information, start and stop dates of the drug of interest, together with any events that have occurred since the patient started the drug under observation.

An event is defined in PEM as: any new diagnosis, reason for referral to a consultant or admission to hospital, any unexpected deterioration (or improvement) in a concurrent illness, any suspected drug reaction, any alteration of clinical importance in laboratory values or any other complaint considered of sufficient importance to enter in patient's notes. The significance of event data is that the reporting doctor is not asked to assess whether there is an association between each event and the drug. The events are then coded into the database using a hierarchical in-house dictionary. This is very similar in structure to the Medical Dictionary for Regulatory Activities (MedDRA), where our lower level terms (LLT) are equivalent to the preferred term used in MedDRA.

The DSRU database currently contains 4.5 million prescriptions and we have in-house over one million green forms from 78 completed PEM studies where the median cohort size is 11 296 (interquartile range 8916 to 13 604). A typical PEM study aims to obtain a cohort of at least 10 000 patients with this it has the power to detect three cases of an event if the incidence rate is between 1 in 1000 and 1 in 2000 with a power of 80%.[4] The completed studies are from various therapeutic classes used in primary care, these are shown in table I where the drugs have been classified into system organ classes according to the WHO's anatomical therapeutic chemical (ATC) classification codes.[5]

Table I. Seventy-eight completed prescription-event monitoring (PEM) studies classified by WHO anatomical therapeutic chemical classification codes

System organ class	Number of drugs with completed PEM studies
Nervous system	21
Cardiovascular	17
General anti-infectives for systemic use	12
Alimentary tract and metabolism	9
Respiratory system	9
Musculo-skeletal system	6
Genitourinary system and sex hormones	4

The collection of event data typically 6 months after the GP prescribed the new medicine enables PEM to perform a postmarketing study as soon as the drug is marketed in England. This data increases the understanding of the effect of the medicine in the 'real world' of everyday clinical practice. Methods of signal generation in PEM include comparison of incidence densities (ID); typically the difference between the ID in the first month of exposure and the subsequent 5 months, in addition to comparisons of 'reasons for withdrawal' and IDs. Furthermore, all events are reviewed by clinical research fellows to identify potential signals as well as looking at events that are suspected to be related to the drug and the most common reasons for stopping. Following the identification of a potential signal the events are followed up. In some instances regulatory authorities or other healthcare professionals can suggest signals. The aim of an automated signal generation system is not to supersede these methods but to enhance them.

The introduction of a new computer system in January 2000 has enabled the development of an automated signal generation system. Two methods have been investigated so far; firstly proportional

reporting ratios (PRRs) and secondly, incidence rate ratios (IRRs). Factors which have been considered include the effect the comparator drugs have on the signal raised, whether looking at only events that occur on treatment is important, the dictionary level at which to perform signal generation.

1. Proportional Reporting Ratios

PRRs compare the proportion that a particular event is of the total number of events reported for a specified drug, with the equivalent proportion in the rest of the database. PRRs are used for signal generation in some spontaneous reporting systems,[6] as they do not require a denominator, i.e. how many people actually took the drug. Although PEM data does have a denominator, PRRs were chosen as the first method to be investigated for an automated system due to their simplicity and ease of application. PRRs are calculated as shown in figure 1. Where the PRR value is greater than 1 the event of interest occurs more frequently in the drug of interest than in the rest of the database. Statistical significance was determined using Yates' correction for the chi-squared (χ^2) calculation based on one degree of freedom. A chi-squared value greater than 4 indicates a p-value of less than 0.05.

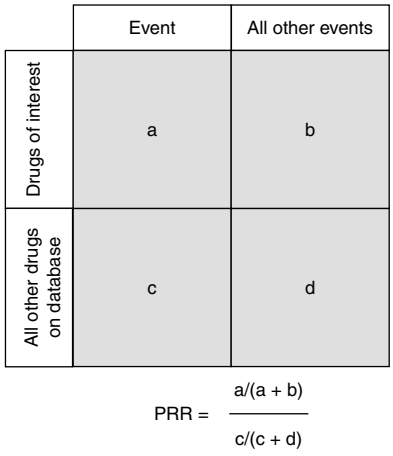

Fig. 1. Calculation of proportional reporting ratios (PRR).

For each drug of interest the PRR versus chi-squared value can be plotted graphically. The symbols indicate the number of reports of an event for the drug of interest, events with less than three reports were excluded from the graph for clarity (figure 2). Potential signals are in the top right quadrant where events have a high PRR (PRR > 1.5), and a high chi-squared value ($\chi^2 > 4$). The drug of interest in figure 2 is montelukast (a leukotriene antagonist licensed for the treatment of asthma), compared with the rest of the database. Many of the potential signals in the top right quadrant are associated with the underlying disease; this is because PEM monitors events and not adverse drug reactions (ADRs). For example, there were 1055 reports of 'steroid short course' (PRR 26, χ^2 25131) for montelukast and this event is unlikely to be a true signal and is more a reflection of the indication. A possible reason that this was identified as a signal is because the number of events of 'steroid short' course is quite rare in the whole database which is likely to be due to there being only four other drugs used for the treatment of asthma in the database.

1.1 The Importance of the Comparator Group

The problem of working with relatively small databases or databases that contain clusters of drug classes, is that these may bias the comparator group. This may be of particular importance in a pharmaceutical company database where the database profile may contain only a few therapeutic classes. The effect is likely to be much more apparent when the database contains event data rather than spontaneous ADR reports, as some events may be related to indication for particular drugs. An example of how changing the comparator group can alter whether an event is a potential signal or not has been illustrated in our database using hallucinations whilst taking tolterodine. Tolterodine is an anticholinergic prescribed for the

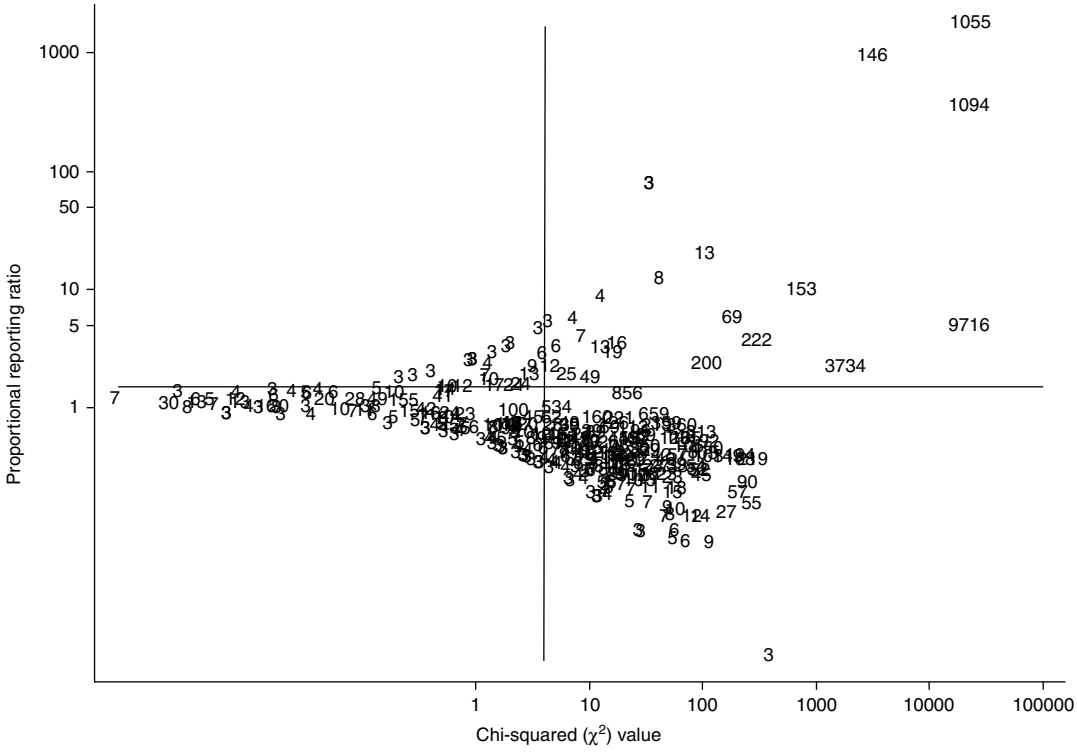

Fig. 2. Plot of proportional reporting ratios (PRRs) versus chi-squared (χ^2) values for montelukast compared with the rest of the database. Symbols indicate the number of reports of an event for the drug of interest, events with less than three reports were excluded. The lines show a PRR = 1.5 and $\chi^2 = 4$, potential signals are in the top right quadrant.

treatment of urinary incontinence. There were 23 reports of hallucinations during tolterodine treatment giving a PRR of 1.1 and χ^2 of 0.13 when compared with the other 77 drugs on our database (see figure 3). In this first comparison a low PRR of 1.1 suggests that hallucination was not a possible signal as it is in the bottom left quadrant of the graph. Subsequently the comparator group was changed to exclude 17 antipsychotic and antidepressants drugs, as some of these drugs have been associated with hallucinations.[7,8] The results of the change in comparator group are shown in figure 4, the PRR for hallucinations increased to 3.1 and a χ^2 27.6, resulting in this being classified as a potential signal (in the top right quadrant of the graph). The example of infrequent hallucinations during tolterodine treatment was used as a positive control since this has already been raised as a potential signal using our data.[9]

A future consideration is whether it will prove useful to remove certain drugs from the comparator group for each specific event where the event is a manifestation of the underlying disease or has already been raised as a drug safety issue. This however will be very labour intensive and time consuming and the question of its value when the aim is to generate signals and not to validate them. A signalling system is aimed to highlight an issue and not to assess the causality and it is important to be aware that the comparator group may have an effect on the results.

PRRs are now being used routinely at the Medicines Control Agency in the UK,[2] they are a useful and simple method that can be applied to event data as well as spontaneous reports of ADRs. There are, however, a few significant differences between PEM and spontaneous reports that do have an effect on the data produced. For PEM, all health-related events are reported by GPs subsequent to prompting after receipt of green form questionnaires, whereas all healthcare professionals in primary and secondary care report suspected ADRs in spontaneous reporting systems. Also, PEM only monitors the first 10 000 to 20 000 people prescribed the drug on NHS prescriptions by GPs, whereas spontaneous reports are sent during the whole postmarketing period. Another difference is that the GP is not asked to establish a causal link between events and drug treatment, whereas spontaneous reporting requires the doctor to have a suspicion that the ADR may be attributable to the drug. Conversely in PEM detailed follow up on events of interest can be obtained and reports often contain better quality data than most spontaneous reporting systems. Lastly, a disadvantage of PEM is that it does not cover all drugs on the market.

The main advantage of performing signal generation with PEM data compared with spontaneous reports is the availability of denominators, these can be split into three types. Firstly, the patient observation period is the study period and is typi-

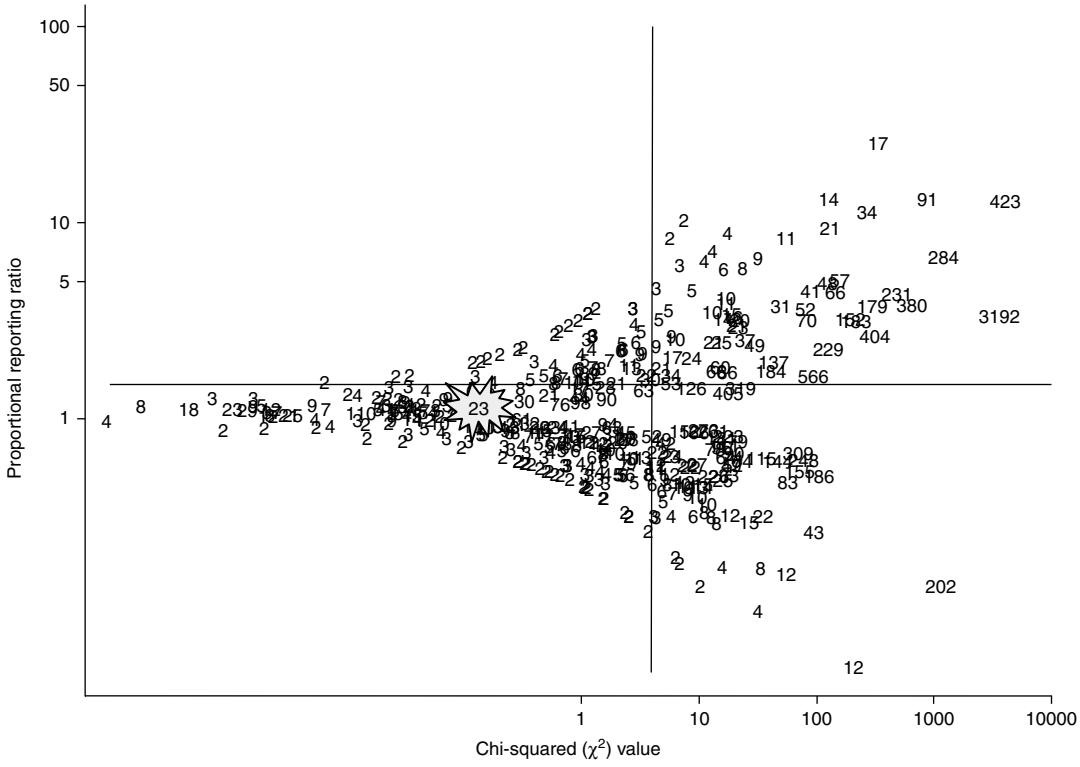

Fig. 3. Plot of proportional reporting ratios (PRRs) versus chi-squared (χ^2) values for tolterodine compared with the other 77 drugs on the prescription-event monitoring database. Symbols indicate the number of reports of an event for the drug of interest, events with less than three reports were excluded. The lines show a PRR = 1.5 and $\chi^2 = 4$, potential signals are in the top right quadrant. The 23 reports of hallucinations are highlighted with a star (PRR 1.1, χ^2 0.13).

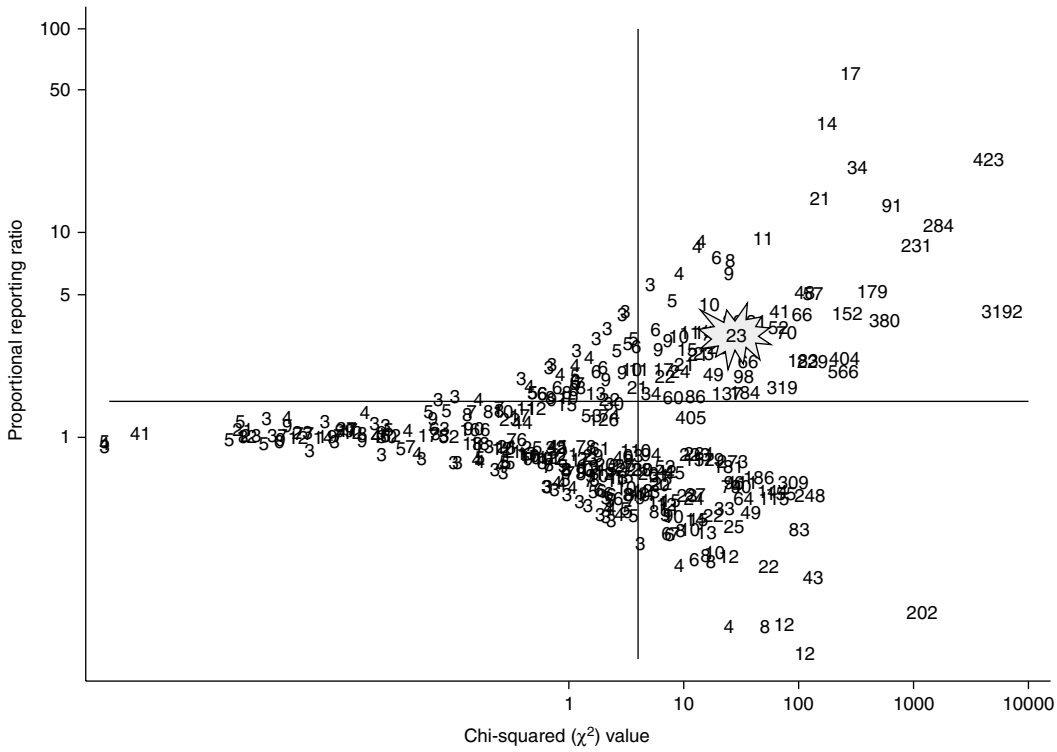

Fig. 4. Plot of proportional reporting ratios (PRRs) versus chi-squared (χ^2) values for tolterodine compared with the 60 of the other 77 drugs on the prescription-event monitoring database (17 antipsychotic and antidepressant drugs were excluded from the comparator group). Symbols indicate the number of reports of an event for the drug of interest, events with less than three reports were excluded. The lines show a PRR = 1.5 and χ^2 = 4, potential signals are in the top right quadrant. The 23 reports of hallucinations are highlighted with a star (PRR 3.1, χ^2 27.6).

cally between 6 to 12 months. Secondly, we can calculate the time each patient was receiving the drug ('patient exposures'). Finally, the patient years at risk (pyar) for each event can also be determined whereby a patient is only at risk of an event whilst receiving the drug or until they have an event. In some cases where the drug has a long half life, a carry over period (often five half lives) may be included where the patient has stopped the drug but they are still at risk of having an event due to the high drug levels still in their bodies. Using these denominators it is possible to take into account time as a factor in signal generation enabling incidence rate ratios (IRRs) to be studied.

2. Incidence Rate Ratios

IRRs compare the incidence rate of a selected event in the drug of interest with the incidence rate for the chosen comparator group. Using the pyar denominator, the incidence rate (λ) for each event was calculated using PEM data for both a drug of interest and the comparator group by dividing the number of new cases of an event in a specified time period (r) by the pyar during the same specified time period. Thus, the IRR is the ratio of the incidence rate in the drug of interest compared with the incidence rate in the comparator group. Statistical significance was by determined using Poisson distribution with a continuity correction on the z test.

IRRs are illustrated in an analogous way to the PRR graphs by plotting IRRs versus the absolute values of the z test (see figure 5). An IRR of greater than 1 means the incidence rate for that event in the drug of interest is higher than that event in the rest of database. The z statistic is equivalent to the square root of the chi-squared value, a z value of greater than 1.6 gives a p-value of less than 0.05. Potential signals are again in the top right quadrant; however, the precise position of the threshold for an IRR value has yet to be established, currently this is in the development stages and is set at 1 so as not to exclude any potential signals. Figure 6 is a plot of lamotrigine, (one of the new generation antiepileptics), when looking at the IRRs after 180 days observation, comparing them with the rest of the database. Two very strong signals represent 12 cases of Stevens-Johnson syndrome (IRR 131, z 19) and 805 events of convulsion (IRR 28, z 113). Using the IRRs and looking at the same events at different time points in the study it is possible to consider time as a factor in signal generation. Figure 6 shows lamotrigine, compared with the rest of the database at 30 days. There are fewer points on the graph, this is probably due there being less events with three or more reports in the first month. These two signals that are clear at 180 days are still evident but the signal is produced from fewer reports; nine reports of Stevens-

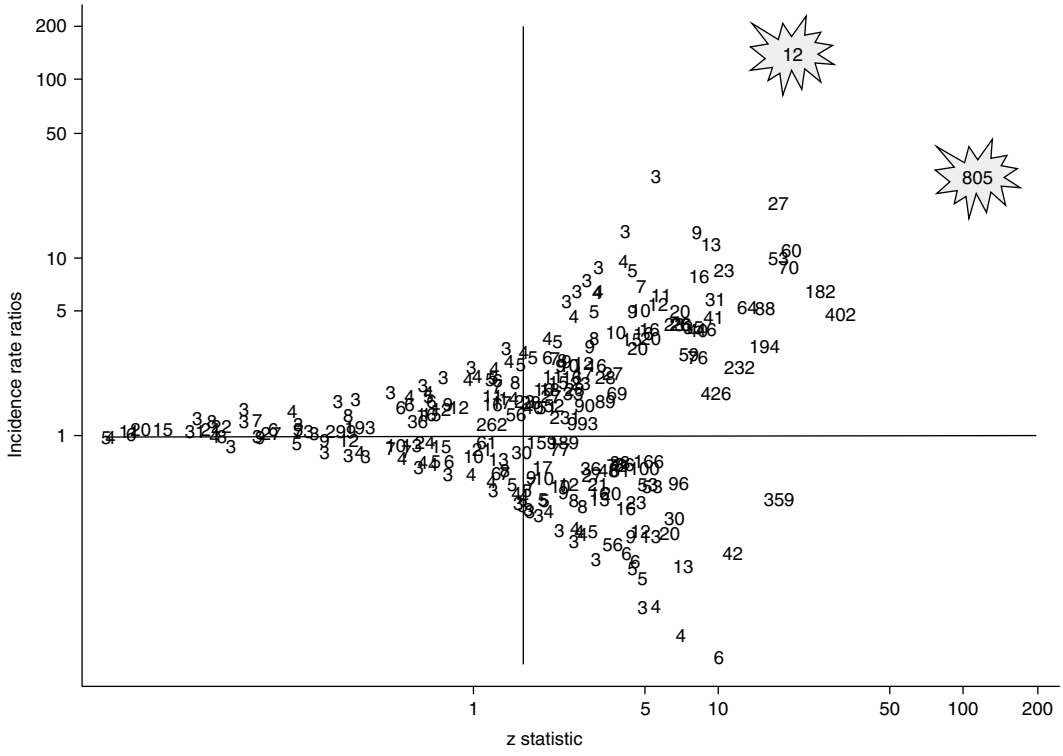

Fig. 5. Plot of incidence rate ratios (IRRs) versus absolute z values after 180 days' observation for lamotrigine compared with the other 77 drugs on the prescription-event monitoring database. Symbols indicate the number of reports of an event for the drug of interest, events with less than three reports were excluded. The lines show an IRR = 1 and z value = 1.6, potential signals are in the top right quadrant. Two very strong signals are highlighted with stars: 12 cases of Stevens-Johnson syndrome (IRR 131, z 19) and 805 events of convulsion (IRR 28, z 113).

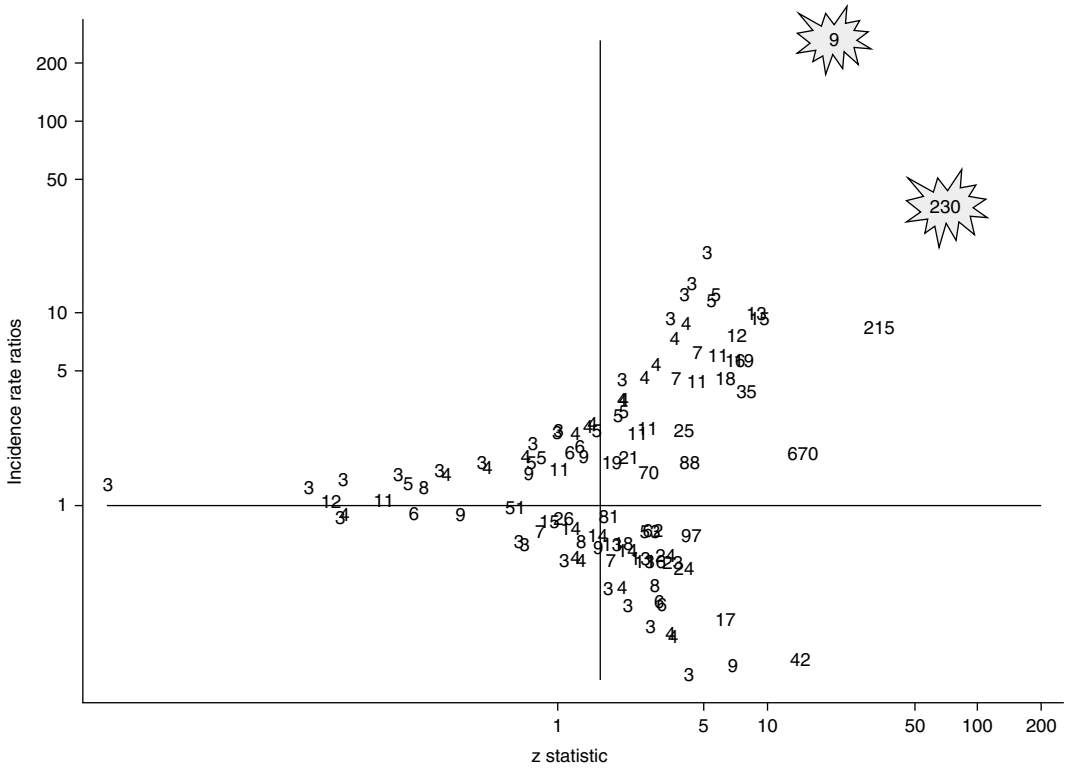

Fig. 6. Plot of incidence rate ratios (IRRs) versus absolute z values after 30 days' observation for lamotrigine compared with the other 77 drugs on the prescription-event monitoring database. Symbols indicate the number of reports of an event for the drug of interest, events with less than three reports were excluded. The lines show an IRR = 1 and z value = 1.6, potential signals are in the top right quadrant. Nine reports of Stevens-Johnson syndrome (IRR 264, z 20) and 230 events of convulsion (IRR 35, z 69) are highlighted.

Johnson syndrome (IRR 264, z 20) and 230 reports of convulsion (IRR 35, z 69).

With the IRRs it is possible to look at the pattern of signals over time. For instance, table II looks at the IRRs for four events in lamotrigine compared with the rest of the database at 30 and 180 days after starting the drug. The IRRs for Stevens-Johnson syndrome, erythema multiforme and general rash are smaller at 180 days than at 30 days after starting lamotrigine. The signals for these three types of rashes get weaker over time because most of the events are reported in the first month. Events like these rashes that have a high incidence in the first month might indicate that the event is associated with starting the drug. Upper respiratory tract infections (RTI) is presented in the same table to highlight that the IRR does not change over time for all events. Upper RTI was specifically chosen as it is an event that is unlikely to be related to the drug, the IRR is 1.6 at both 30 and 180 days sug-

Table II. Four incidence rate ratios (IRRs) for lamotrigine at 30 and 180 days

Event	IRRs	
	30 days	180 days
Stevens-Johnson syndrome	265	131
Erythema multiforme	19.6	6.9
Rash	8.0	4.4
Upper respiratory tract infection	1.6	1.6

gesting that this event is not related to time on the drug. The IRR of 1.6 suggests that upper RTI occur slightly more frequently in lamotrigine compared with all the other drugs in our database. A plausible explanation for this is that the average age of the lamotrigine cohort is lower than the rest of the database as it contains a higher proportion of children.

3. Selecting the Appropriate Dictionary Level

Potential signals can not only be lost because of choice of the comparator group but they can also be missed through problems with the terminology. The level in the dictionary at which signal generation is performed is important. The dictionary at the DSRU is an in-house dictionary that is similar in structure to MedDRA, our LLT are equivalent to the preferred terms in MedDRA, many LLTs can be grouped into a higher level term (HLT) in a hierarchical fashion. For example the LLT of 'colitis' is grouped together with 'pseudomembranous colitis', 'ulcerative colitis', 'perforated colon', 'Crohn's disease', 'diverticulitis', 'diverticulosis', 'diverticulum', 'improved diverticulosis' under the HLT of 'inflammatory disease colon'. The new procedures at the DSRU enable signal generation to be performed on both LLT and HLT. The disadvantage of the LLT is that there are 1663 terms compared with 1143 HLT. In some cases there are only a small number of reports at each LLT; however, when the terms are grouped into a HLT of a broader clinical category the number of events increases. Therefore at the LLT level it is harder to gain statistical significance due to the small numbers. Together with excluding event with less than three cases it is possible to miss potential signals that appear when grouped under a HLT. However, the same can be true where the HLT is too broad and contains many LLTs a potential signal can be missed. An example of this was illustrated in our database with one of the nonsteroidal anti-inflammatory drugs (NSAIDs) and a potential signal of colitis. Both PRRs and IRRs at HLT and LLT were performed on the NSAID compared with the other 77 drugs. The HLT of 'inflammatory disease colon' is raised as a very weak signal with an IRR of 1.7 (z value 2.5) and a PRR of 1.8 (χ^2 11.7), whereas the more specific LLT 'colitis' is a strong signal with an IRR 5.7 (z value 5.3) and a PRR 3.8 (χ^2 18.2). This suggests that a step down approach may be an appropriate way of tackling signal generation whereby HLT are investigated first and then the more specific LLT.

4. Conclusion

Signal generation is a fishing expedition or as previously likened to like finding a needle in a haystack.[10] Automated signal generation is a way of producing a hypothesis of a potential link between a drug and an event, but this does have limitations that should not be ignored. Foremost, it does not assess the causality of an association between an event and a drug, but merely highlights that the frequency of the event is higher than in the comparator group, thus requiring reviewing of the individual cases and possibly further studies to confirm or dispute the proposed hypothesis. Another limitation is where there are two or fewer events as it is very hard to establish a true signal on such small numbers. However, if the event is rare and serious, where a high proportion of such events are due to drugs e.g. aplastic anaemia, automated signal generation is not required, the research fellow at the DSRU responsible for the drug identifies any serious events when they first assess each questionnaire returned.

Consideration should be taken into the dictionary level used and whether it is appropriate to perform signal generation on more than one level in the dictionary as is appropriate with the DSRU dictionary. The structure of the dictionary is likely to be of great importance and the role of MedDRA in signal generation should be studied extensively. This paper has highlighted the potential pitfalls of using the rest of a database as the comparator group, it is likely that the selection of the compa-

rator group will depend upon drug profile of the database being used.

This is a methodology that we have performed retrospectively, and we are now moving onto current studies. Other future ambitions are to use gender-specific denominators for gender-related events such as gynaecomastia. Currently, it is difficult to consider each individuals time to the event, this could be achieved by performing survival analysis as part of routine signal generation. Whereby the survival experience of the patients for a particular event in the drug of interest is compared with that in the comparator drugs, the difference can be tested using a statistical analysis such as a Log rank test.

Acknowledgements

We thank Professor Stephen Evans for his invaluable advice on the statistical analysis. Along with the AstraZeneca foundation for the research grant for Dr Emma Heeley, we are also very grateful to the general practitioners who voluntarily take part in prescription-event monitoring studies.

References

1. 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 (4): 315-21
2. Evans SJW, Waller PC, Davis S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. Pharmacoepidemiol Drug Saf 2001; 10: 483-6
3. Mann RD. Prescription-event monitoring: recent progress and future horizons. Br J Clin Pharmacol 1998; 46 (3): 195-201
4. Machin D, Campbell MJ, Fayers PM, et al. Sample size tables for clinical studies. 2nd ed. Boston; Blackwell Science, 1997
5. Ronning M, Blix HS, Harbo BT, et al. Different versions of the anatomical therapeutic chemical classification system and the defined daily dose: are drug utilisation data comparable? Eur J Clin Pharmacol 2000; 56 (9-10): 723-7
6. Evans SJW, Waller P, Davis S. Proportional reporting ratios: the uses of epidemiological methods for signal generation [abstract no. 55]. Pharmacoepidemiol Drug Saf 1998; 7: S79
7. Bourgeois JA, Thomas D, Johansen T, et al. Visual hallucinations associated with fluoxetine and sertraline. J Clin Psychopharmacol 1998; 18 (6): 482-3
8. Coryell W. Psychotic depression. J Clin Psychiatry 1996; 57 Suppl. 3: 27-31
9. Layton D, Pearce GL, Shakir SA. Safety profile of tolterodine as used in general practice in England: results of prescription-event monitoring. Drug Saf 2001; 24 (9): 703-13
10. Edwards IR. Adverse drug reactions: finding the needle in the haystack [editorial]. BMJ 1997; 315 (7107): 500

Correspondence and offprints: Dr *Emma Heeley*, Drug Safety Research Unit, Bursledon Hall, Blundell Lane, Southampton, SO31 1AA, UK.
E-mail: emma.heeley@dsru.org