ORIGINAL RESEARCH ARTICLE

Drug Saf 2012; 35 (7): 599-610
0114-5916/12/0007-0599/$49.95/0

A Signal Detection Method to Detect Adverse Drug Reactions Using a Parametric Time-to-Event Model in Simulated Cohort Data

Victoria R. Cornelius,[1] *Odile Sauzet*[2] and *Stephen J.W. Evans*[3]

1 Primary Care and Public Health Sciences, King's College London, London, UK
2 Epidemiologie & International Public Health, Universität Bielefeld, Bielefeld, Germany
3 The London School of Hygiene and Tropical Medicine, London, UK

Abstract

Background: Current quantitative signal detection methods have been primarily developed for the purpose of detecting signals from spontaneous reports. These methods are not always appropriate for cohort data. More recently, parametric time-to-event models have been proposed to model hazard functions with the ultimate aim of detecting adverse drug reactions (ADRs). The rate of occurrence of ADRs after starting a drug will depend upon the causal mechanism and therefore will often vary with time, in contrast to events not associated with the drug, which will tend to occur at a constant background rate. After starting treatment, the onset of ADRs will be rapid for some but delayed for others. A non-constant rate over time may indicate a drug-event relationship.

Objective: The aim of this study was to propose a simple test to detect signals of ADRs in cohort data and to investigate the power of this test using simulated data. A signal detection tool using the proposed test to improve the power of detection is also described.

Method: In order to test for a non-constant hazard (rate of occurrence), the hazard function was estimated using the model shape parameter for the Weibull function. If the shape parameter was found to be significantly different ($p < 0.05$) from the value one (the value for a constant hazard) a signal was raised. Simulation of background event rates used were 1%, 5% and 10% of the cohort size. The ADR rate was varied in proportion to the background rate; a 10%, 20% and 50% increase in the background rate was explored. The time of occurrence of the ADR will dictate the shape of the hazard function, therefore the ability of the model to detect a signal depending when the highest risk for ADR was also explored. The power of the test was investigated by simulation.

Results: The Weibull Shape Parameter (WSP) test was most powerful at detecting signals that occur shortly after starting treatment. These preliminary simulations had low power when the underlying hazard function was symmetrical (e.g. when ADRs occurred in the middle of the study period). The power of the test was improved by censoring the data as this broke the symmetry of the hazard function. A tool that censored the data at regular intervals and repeated the WSP test was found to correctly detect ADR or no ADR around 90% of the time when the sample size was at least 5000.

Conclusion: The WSP test is simple to implement using standard statistical software, and can be used to detect non-constant hazards over time in order to raise signals of time-dependent ADRs. When there is no pre-specified event of interest or the time of the ADR is uncertain, the WSP tool should be used instead of the WSP test. These methods do not require any external data for comparative purposes and thus can be implemented in a single cohort of participants exposed to a drug.

Background

Monitoring the safety of newly marketed medicines has predominantly been achieved using spontaneous reporting systems. In the UK, a Yellow Card system is in place where healthcare professionals and patients are encouraged to report suspected adverse drug reactions (ADRs);[1] similar systems are in place for other countries. Spontaneous reports contain limited information on subjects with suspected ADRs. They collect information such as the severity of the ADR, the outcome of the ADR (e.g. resolved, lasting effects, death), limited patient characteristics data, the date and length of ADR and date the drug was started. However, such data are frequently missing the details and it is often the case that no dates are available. In addition, when using such systems it is not possible to determine the total number of subjects who are treated with a drug at a particular time and therefore it is not possible to estimate statistics such as the incidence density or risk. Spontaneous reporting systems rely on the clinical professional or patient involved identifying a potential causal relationship between the observed adverse event and the drug in order to report it; this link can be more difficult to identify as the duration following the onset of the drug increases. The limitations of spontaneous reporting have been well documented.[2] Pharmaceutical companies, academics and regulators have developed a suite of automated signal detection methods in order to help them to identify ADRs from the database of all suspected ADRs. These methods are primarily based on the principle of 'disproportionality' which involves comparing the observed number of reports for a specific drug-adverse event combination to the expected number, which is derived from the total database of spontaneous reports.[3]

However, sources of longitudinal observational data on cohorts of patients receiving a drug are increasingly available. Pharmaceutical companies are required to perform postmarketing studies to monitor the safety of newly marketed medicines, and these are often observational in design, such as the Prescription Event Monitoring studies in the UK.[4] In addition to studies performed by industry, routinely collected clinical information is available through longitudinal population base databases (electronic health records [EHRs]) such as those included in the EU-ADR project (Exploring and Understanding Adverse Drug Reactions by integrative mining of clinical records and biomedical knowledge).[5] New developments in study design for pharmacovigilance have also resulted in proposed systems that collect data longitudinally.[6] These longitu-

dinal datasets will contain information on all adverse events that happen to a participant when taking a drug and not just adverse events suspected to be ADRs. The studies performed using such data can be either prospective or retrospective in design. The types of adverse events included can either be pre-specified or include all adverse events that happen to a participant during the observation period. For this type of study, the date of each adverse event and the treatment duration is known for each subject (whether or not they have an adverse event); as a result, statistics such as the incidence density can be calculated. These statistics are often compared across cohorts of patients taking the same class of drug despite the difficulties with making such a comparison. The cohorts are unlikely to be directly comparable due to issues such as selective prescribing and channelling. Other comparisons made using this type of data include choosing an arbitrary timepoint and comparing the incidence density between the two time periods either side of this timepoint by calculating differences; however, the choice of the cut value is problematic and the power of this method will be partially dependent upon this selected cut value.

A drawback of utilizing established signal detection methods (that have been developed for spontaneous reporting) on cohort data is that a large number of similarly designed 'cohorts' would be required to act as a reference database in order to derive an expected value that could then be compared to the observed value in the cohort of interest and this may not always be available. In addition, the time-to-event data (which is available for cohort data) will not be utilized when using these methods. A further method, which requires a pre-specified period defined to be a 'high-risk' period, uses sequential methodology applied to the self-controlled case series method.[7] The advantage of this approach is that it requires cases only (participants with events). Recently, an adaptation of the Gamma Poisson Shrinker (GPS) for longitudinal data (LGPS), which enables the time the participant is on a drug to be taken into account, has been proposed.[8]

In this paper we propose a simple test for detecting signals of ADRs in cohort data (of patients exposed to a drug) that takes into account the duration between start of treatment and the event. The test is based on fitting a Weibull model to the time-to-event data in order to detect a variation in hazard of an event due to ADRs occurring during a restricted period of time. We investigate how this test performs (power) on simulated data according to various background rates of events (not due to the drug). We then propose a detection tool for ADRs based on this test.

Detecting a Non-Constant Hazard over Time

The rate of occurrence of time-dependent ADRs after starting a drug will generally vary with time. The onset of some ADRs will be rapid after starting treatment, whilst for others the onset may be delayed. So for a given drug, the time in which ADRs occur is restricted, unlike the occurrence of events not due to the drug; these can happen at any stage during the study observation period. A constant hazard over time can be consistent with observing a background event rate, whereas a non-constant hazard over time may indicate a drug-event relationship.

The hazard function is the (limiting) probability that an event occurs in a given (small) time period, conditional on the patient surviving until the start of that time period. If none of the events observed in the population are due to the drug, then the hazard function for the event of interest on average will be constant over time.

Method

Model for the Simulated Data

Time-to-event data was simulated for each individual in the cohort. The time-to-event was defined as the duration between the start of medication and the onset of the adverse event (which will be either an ADR or a background event), the time the patient stopped medication or the end of the study observation period, whichever was the soonest. (Note that when the participant is no longer being 'observed' for the event of interest, then the participant status becomes 'censored' at this time).

The times of adverse events not associated with the drug were generated using a uniform random sample over the study time period. The number of such events was determined by the cohort size and the background rate. The times of events due to the drug (ADRs) were generated using a normal distribution centred on the middle of a selected month, with a standard deviation of 0.25, which on average will assign 95% of ADRs occurring during the given month (note that ADRs with simulated occurrence time before or after the defined study period were treated as 'unobserved' events). This month was then defined to be the month of highest risk. The impact of assuming a normal distribution was explored by repeating the simulations assuming a log-normal and then a uniform distribution for the times of the ADRs (results not shown). The number of ADRs was defined by the cohort size, the background rate over the observation period, and the risk of ADRs due to the drug as a proportion of the background rate. We use the term 'month' as a time unit which, in practice, can be any fraction of the observation period.

Signal Detection Test for Adverse Drug Reaction

The signal detection test is based on the identification of an increased value of the hazard function during the month of highest risk for the occurrence of ADRs. If no ADRs take place in the observed population, then the hazard of an event occurring will be constant during the study period. However, if ADRs occur in a limited period of time during the follow-up period then the hazard will not be constant.

Distribution of event times that occur at a constant rate in a cohort is called the exponential distribution. Survival curves plotted on a log scale where the rate (hazard) is constant form straight lines, with a slope dependent on the constant rate. If the rate is not constant, but, say, is high initially then reduces over time, then one of the distributions that can describe such data is the Weibull distribution. To describe this distribution requires two parameters, one of which essentially measures how far the distribution is from a non-constant hazard. The exponential distribution is a special case of the Weibull and it is possible to test whether there is evidence for a non-constant hazard rate. The baseline hazard (h) function for the Weibull model has the form given in equation 1.

$$h(t) = \lambda t^{\lambda - 1} \exp(\beta_0) \qquad \text{(Eq. 1)}$$

where t = time, β_0 is a constant and λ is the shape parameter. If the shape parameter is equal to one, then it can be seen from equation 1 that the hazard function reduces to a constant. If the shape parameter is different from one then the hazard will not be constant over time.

Weibull Shape Parameter (WSP) Test

A significance test for a shape parameter equal to one (the value of a constant hazard) is performed by fitting a Weibull model to the data. If the estimate of the Weibull Shape Parameter (WSP) λ has an associated p-value lower than 0.05 then the shape parameter has been found to be significantly different from the value one, the hazard is considered to be non-constant and a signal is raised for a possible ADR.

Assessment of Power and Development of a Signal Detection Tool

This work was developed in two phases. The first phase assessed the power of the WSP test to detect the ADRs generated in the simulated dataset. In the second phase, a tool that is based on the detection test is described.

A variety of scenarios were generated for a 6-month observation period in which the size of the cohort, the background rate of events, the risk of the ADR, and the time at which the ADRs took place after starting treatment were altered. The subjects' time did not include censored information (i.e. all subjects were assumed to have been observed and on medication for the full 6 months). The background event risk (events unrelated to treatment) was defined in relation to the size of the cohort. Background risk of events explored were 1%, 5% and 10% of the cohort size for a 6-month observation time. The ADR risk varied in proportion to the background risk;

a 10%, 25% and 50% increase in the background risk was explored. The three time periods considered were shortly after starting treatment (month 1), medium term (month 3) and long term (month 6). For ADRs occurring in the middle of the study period, we also investigated the power of the test when the observations were censored at the end of the third month. The power of the test as well as the sample size required to attain 80% and 90% power were obtained for all the scenarios.

An estimate of the power of the WSP test was obtained by computing the p-value for 20 000 simulated datasets for each scenario (this number of simulations were chosen so as to provide stable results for the power). The sample sizes required for 80% and 90% power were obtained by simulating 30 000 cohorts incrementing the frequency of the sample size by ten until the test was able to detect ADRs in 80% or 90% of cases. Graphs are provided to show the results of all the simulated scenarios and a table of the sample size required to achieve an 80% or 90% power for simulated scenarios is given.

In the view that the above test was performing better when the highest risk for ADRs occurred either in the first or last month of observation (see Results and Discussion sections) the following tool to detect the presence of ADRs in a dataset was derived: censor the data at regular intervals throughout the study period (regardless of when the distribution of the observed events or expectation of ADR timing due to causal mechanism as this will not always be known) and then perform the detection test for each of these censored datasets (e.g. for this study the data were censored in turn at months 2, 3, 4, 5 and 6) . A signal is then raised if the test is positive ($p < 0.05$) in any one of the datasets.

WSP Detection Tool

For pre-defined regular time interval t_i, where $i = 1$ *to* n (e.g. end of each month):

- Create a dataset with subject data censored at time t_i (all subjects who have observations beyond t_i will have their observation period reduced to t_i. Only adverse events that happen prior to t_i will be included; if adverse events happen after t_i then the event status is changed to 'no event').
- Perform the WSP signal detection test.
- Repeat for $i = 1$ *to* n.

If one or more of the p-values for the t_i censored periods obtained is lower than 0.05 then a signal for the possible presence of ADRs is raised.

We investigated the ability of the WSP detection tool to accurately detect the presence of ADRs for a variety of sample sizes, background rates and increase in rate due to ADRs. The simulations were created as described above, but instead of pre-defining the month in which the ADR distribution was centred, we randomly selected a month between 1 and 12 in which the ADRs occurred. This meant that there would be ADRs in approximately one-half of the simulated datasets and when ADRs were present they could be in any one of the 6 months.

We performed 30 000 simulated datasets for each scenario, and reported the percentage of cases when the results are accurate (ADRs correctly detected or no ADRs correctly detected), ADRs wrongly detected (false positive) and no detection of ADRs when they were present (false negative) as well as sensitivity and specificity of the detection tool.

Results

The power of the WSP test to detect a signal when a true signal is present can be seen for each simulated scenario in figures 1–4. The figures show the power achieved by the test (percentage of cases for which ADRs have been detected) against cohort size, for each background rate and percentage increase of the background rate (ADRs) centred in month 1 (figures 1a–c), month 3 (figures 2a–c), month 3 with data censored at the end of the third month (figures 3a–c) and month 6 (figures 4a–c).

Table I shows the sample size required to achieve 80% and 90% power to detect ADRs in a given month (months 1 and 6, and month 3 with data censored at the end of month 3) for the different scenarios, together with the corresponding number of background events and ADRs. When

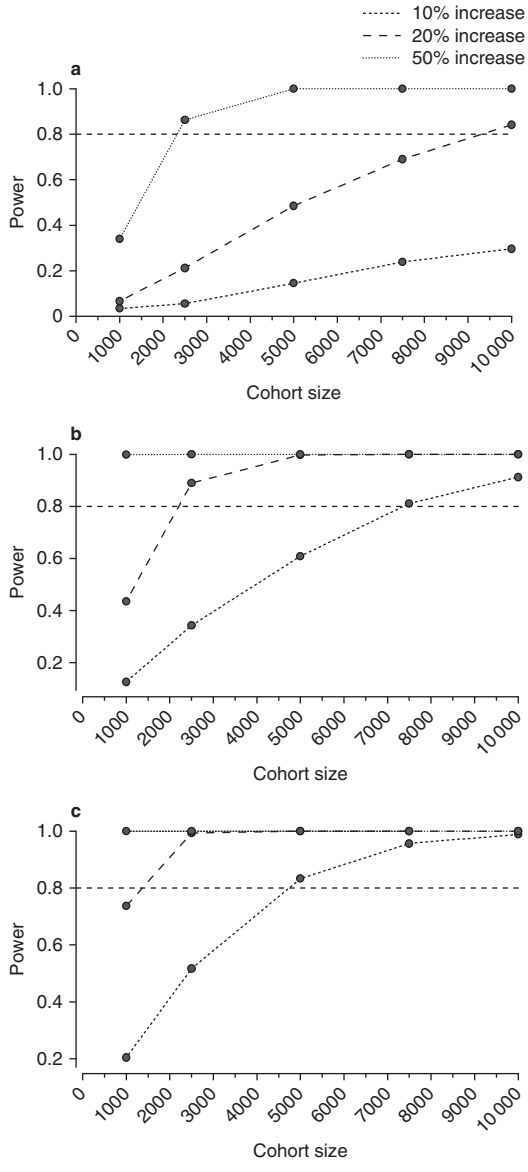

Fig. 1. Power of the Weibull Shape Parameter test by sample size when adverse drug reactions occur shortly after the start of treatment (month 1) for 10%, 20% and 50% increase in risk compared with the background event rate of (**a**) 1%, (**b**) 5% and (**c**) 10%.

the background rate and increase in the background rate due to ADRs was very low, the sample size required to achieve 80% and 90% was too large to be achievable; this is represented by a dash in the table.

ADRs with the highest risk of occurrence in month 1 are always well detected when the rate of ADRs is at least 50% of the background rate. If the rate of ADRs is 20% of the background rate this will depend on the cohort size but if the cohort size is large enough, ADRs will be detected in more than 80% of cases (figures 1a–c). The same is also true when the rate of ADRs is 10% of the background rate, except for the situation when the background rate is 1% (figure 1a), where the power remains very low. The exact sample size required to detect a signal with 80% and 90% power can be seen in table I.

ADRs with the highest risk of occurrence in month 6 are always well detected for a large enough cohort when the rate of ADRs is at least 50% of the background rate (figures 4a–c). When the background rate is 1% a signal is not well detected when the ADR is 10% or 20% above this background rate (figure 4a). When the ADR rate is at least 20% of the background rate and a background rate is at least 5%, ADRs will be well detected if the cohort size is large enough (figure 4b). The WSP test only detects an ADR rate of 10% with at least 80% power when the background rate is at least 10% (figure 4c compared with figures 4a and 4b).

The test for ADRs with the highest risk of occurrence in month 3 lacked power in all but the most extreme scenario of a rate of 50% of the background rate and a background rate of 10% (figures 2a–c). However, if the observations are censored at the end of the third month the situation is very similar to ADRs with the highest risk of occurrence in month 6 with cohort sizes being slightly higher to achieve a similar power (figures 3a–c).

It can be seen from table I that the sample size required to detect signals where the increase in background rate due to ADRs is 20% and 50% is substantially less that that for a 10% increase. For example, the actual number of events (background events plus ADRs) required to detect a 20% increase in background events (due to ADRs) in the first month after starting treatment with 80% power, no more than 140 events are necessary regardless of cohort size. Whereas, for a 10% increase, the number of events (background events

plus ADRs) has to be at least 400. This number increases substantially with background rate. This feature is common regardless of when the ADRs occur.

Table II presents the results for the reliability of the WSP detection tool, which consists of censoring data at the end of months 2, 3, 4, 5 and 6, and performing the WSP test for each situation. The table gives, for each scenario, the percentage of correct answers (where the tool correctly detected a signal or correctly did not detect a signal), the percentage of cases when an ADR has been incorrectly detected (false positive) and the percentage of cases where an ADR was missed (false negative) as well as the sensitivity and specificity.

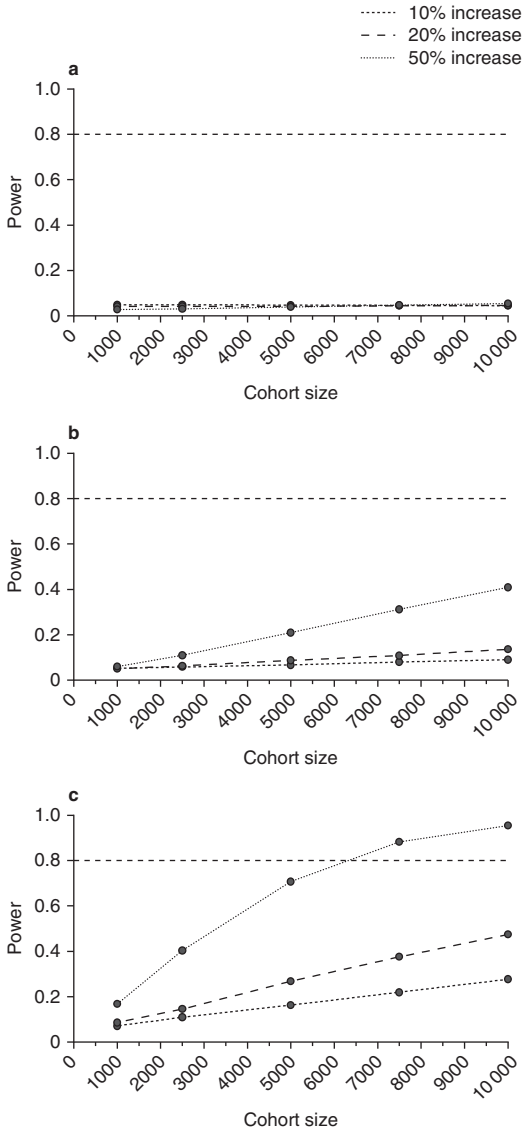

Fig. 2. Power of the Weibull Shape Parameter test by sample size when adverse drug reactions occur medium term (month 3) for 10%, 20% and 50% increase in risk compared with the background event rate of (**a**) 1%, (**b**) 5% and (**c**) 10%.

When the background rate is 1% and the expected increase in the background rate due to ADRs is 20% or less, the tool is not useful at detecting a signal. However, if the expected increase is at least 50%, the WSP tool will be able to correctly detect the presence or absence of ADRs in 93% of the time (for a sample size of 10 000) with no missed signals (i.e. no false negative results), giving a specificity of 0.87 and a sensitivity of 1.

When the background rate was 5% or more, the tool was able to detect correctly the presence or absence of ADRs in a high percentage of cases and the false negative rate was low or zero (provided the sample size was at least 1000). In most cases the only errors were false positives. The background rate will affect the percentage of false positives, whereas the rate of the ADRs will affect the proportion of false negatives. The specificity will always be over 0.8 and the sensitivity increases with the percentage increase due to ADRs.

Discussion

This study was designed to explore the power of the WSP test to detect ADRs that increased the risk by at least 10% above the number of events expected not due to the drug. This signal detection test differs from signal detection methods that have been developed for spontaneous reporting systems as it takes into account the time at which events occur (and not just whether the event occurred or not). Another method that takes account of time is a modification of the GPS called LGPS,[8] which is based on the principle of disproportionally and, as a result (similar to the signal detection methods for spontaneous reporting systems), the analysis needs to be

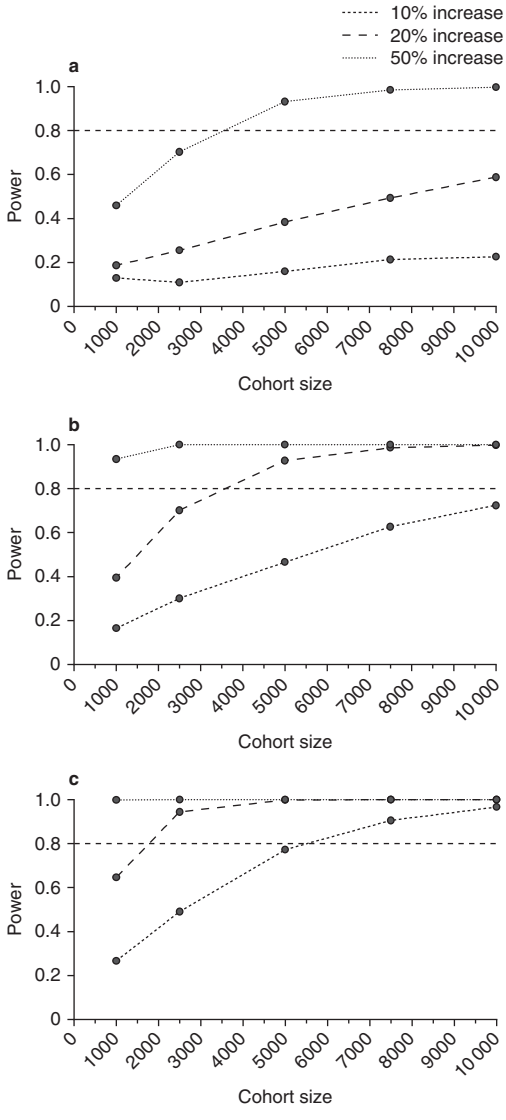

Fig. 3. Power of the Weibull Shape Parameter test by sample size when adverse drug reactions occur medium term (month 3) when data censored at the end of month 3, for 10%, 20% and 50% increase in risk compared with the background event rate of (**a**) 1%, (**b**) 5% and (**c**) 10%.

performed on a database that contains participants who are not exposed to the drug of interest. The database will need to be large enough such that the reference (unexposed) population are suitably diverse and that a particular disease or drug is not overly influential in the calculation. Our method can be applied to single-drug cohort data where there is no obvious comparator. Previous methods of signal detection for single-drug cohort data have included comparing the incidence densities between two time periods by calculating the difference between the incidence density and the 95% confidence interval for this difference. The limitation of this method is that the observation period has to be categorized into two arbitrary intervals, categorization of continuous time data will reduce power to detect signals. An advantage of our method is that there is no need to pre-define time periods of interest in which to test. The time to event data for spontaneous reports is often of poor quality and only offers information on people who have had suspected adverse drug reactions (although the method might be able to be applied there when data are complete). EHRs offer a new and valuable recourse for signal detection. The use of traditional signal detection methods on EHR would not fully use this data and use of a time-to-event approach should provide added precision.

The WSP method provides a new tool for signal detection on prospectively collected single-drug cohort data and routinely collected EHR data. It has the potential to reduce the time to identify unknown ADRs, and to offer a simple signal detection process. Postmarketing drug observational studies (such as those performed by pharmaceutical companies as part of their risk management plan) using Prescription Event Monitoring currently may have no obvious comparator, and usual methods used to identify ADRs on this type of data are likely to be underpowered. This method can be employed in any routinely collected (EHR) data and it does not rely on the clinician or the patient to make a causal link (unlike spontaneous reports) as all 'events' recorded in patient notes can be screened. The use of this method in EHR data needs to be evaluated further as the impact of issues with reporting and recording bias, coding of events and informative censoring need to be understood. It would also be possible to use a modification of the method in spontaneously reported data, where the comparator would be data from events other than the one of interest. The clinical relevance is that the

utilization of the extra information derived from the time pattern may allow for better properties of the signal detection process. Obviously, this suggestion does need much more extensive checking in real data to see if this is the case.

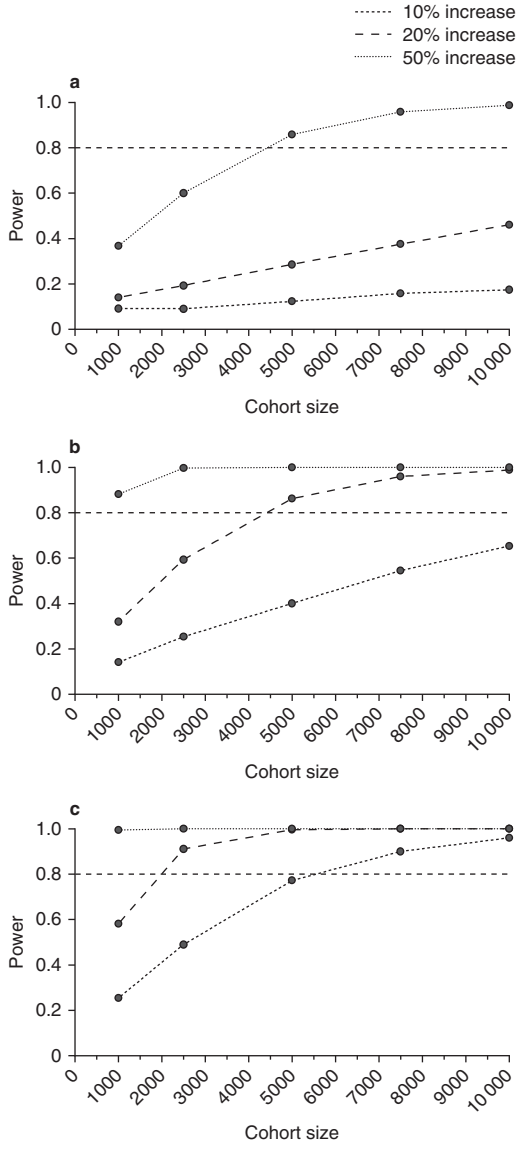

Fig. 4. Power of the Weibull Shape Parameter test by sample size when adverse drug reactions occur long term (month 6) for 10%, 20% and 50% increase in risk compared with the background event rate of (**a**) 1%, (**b**) 5% and (**c**) 10%.

Other authors have suggested using time-to-event modelling for suspected ADRs collected using spontaneous reporting systems in order to determine whether the hazard function is associated with the underlying mechanism of toxicity for a drug.[9] This is somewhat different from what we have proposed in this paper. Their interest is in modelling the hazard function. This requires fitting several models and then determining which model is the best fit to the data in order to obtain a description of the empirical hazard function, this will not necessarily determine whether the hazard function is significantly changing over time. They also perform this using information on subjects with suspected ADRs only, so no information about the subjects who do not have an event is included. In this study, the Weibull model is not used as a prediction tool but as a detection tool for all subjects in a cohort (and not just subjects with an event). Our aim was to achieve adequacy of detection and not model fit.

The preliminary simulations performed in this study show that the signal detection test was most powerful when the ADR occurred shortly after starting treatment; it also had high power when the ADR had highest risk of occurrence at the end of the study observation period. The graphs of the power of the WSP test in figures 1–4 enable readers to extract information regarding sample size and power. For example, when looking at the solid line on figure 1, it can be seen that when the background rate of an event is 1%, a cohort size of 2500 subjects would have at least 80% power to detect a 50% increase in risk of the event. (Cohort size n = 2500, background risk 1% [25 events], ADR risk 50% [12–13 events].)

Table I presents the sample size required to achieve 80% and 90% power for the test to detect a signal for a variety of scenarios. Whilst it was not possible to compute all required scenarios, it is possible to gain an indication of the sample size required to detect a certain increase in percentage risk of an ADR for a given background risk. For example, 2110 participants (or 31 events) would be required to detect a 50% increase in a 1% background rate where ADRs have the highest risk of occurrence within the first month of treatment with 80% power (i.e. 21 events unrelated to

Table I. Sample sizes in terms of participants and number of events required for 80% and 90% power to detect specified risk of adverse drug reactions by month of occurrence using the Weibull Shape Parameter test. Results are based on 30 000 simulations

Percentage of background events	10% increase		20% increase		50% increase	
	80%	90%	80%	90%	80%	90%
Adverse events in the first month						
1%	–	–	9260	–	2110	2510
Number of background events (adverse events)			92 (18)		21 (10)	25 (12)
5%	7300	9500	1950	2550	<1000	<1000
Number of background events (adverse events)	365 (36)	475 (47)	94 (18)	125 (25)	<50 (<25)	<50 (<25)
10%	4550	6050	1130	1430	<1000	<1000
Number of background events (adverse events)	455 (45)	605 (60)	113 (22)	143 (28)	<100 (<50)	<100 (<50)
Adverse events in the sixth month						
1%	–	–	–	–	4110	5710
Number of background events (adverse events)					41 (20)	57 (28)
5%	–	–	4050	5560	<1000	1070
Number of background events (adverse events)			200 (40)	275 (55)	<50 (<25)	53 (26)
10%	5380	7380	1780	2380	<1000	<1000
Number of background events (adverse events)	538 (53)	738 (73)	178 (34)	238 (46)	<100 (<50)	<100 (<50)
Adverse events in the third month (data censored at the end of the third month)						
1%	–	–	–	–	3100	4300
Number of background events (adverse events)					31 (15)	43 (21)
5%	–	–	3250	4450	<1000	<1000
Number of background events (adverse events)			162 (32)	222 (44)	<50 (<25)	<50 (<25)
10%	5350	7290	1530	2080	<1000	<1000
Number of background events (adverse events)	535 (53)	729 (72)	153 (30)	208 (40)	<100 (<50)	<100 (<50)

– indicates the sample size is too large to be achievable.

drug with an additional 10 ADRs expected due to the drug). The table also shows that there are situations where it is not possible to detect ADRs using this method. For example, it is not possible to detect a 10% increase when the expected background risk of an event is 1% with 80% (and clearly 90%) power regardless of when the anticipated ADR would occur with any reasonable cohort size.

When the true hazard function presents some symmetry (for example when ADRs occur in the middle of the observation period), the power of the test is dramatically reduced (figures 2a–c). This is due to the limitation of the Weibull model which can only allow the hazard function to monotonically increase or monotonically decrease, whilst this assumption poses little problem when we have an asymmetrical non-monotonic function; fitting a Weibull model to a symmetric hazard function results in an estimate of a constant hazard over time. However, by censoring the data at the end of the ADR period the symmetry of the hazard function is broken and the test is able to detect the increased hazard and is therefore more powerful, even though information was lost by truncating the observation period. Whilst the time periods in this study are referred to as 'months' the actual units are arbitrary.

In practice we do not always know or understand the toxic mechanism for the ADR and therefore we will not know the expected time of the ADR. In addition, it is desirable to have a tool that has the ability to detect a signal regardless of when that signal occurs. As a consequence we proposed the WSP tool, where the test is performed repeatedly by censoring the data at a number of regular intervals throughout the study observation period. High reliability (over 80% with the correct answer) was achieved for all scenarios provided the sample size was large enough, except

for when the background rate and ADR rate was low (1% and 10%, respectively). In this situation the tool did not seem to be able to distinguish between the sparse background events (not associated with the drug) and true ADRs, as the background events would not necessarily have followed a uniform distribution due to the small number of events. It seems that false negative signals (failing to detect existing ADRs) are related to the background rates and they increase when the background rate decreases. False positive signals (when the tool detects a signal when no ADRs are present) are related to the rate of ADRs. With a large enough sample size, very few true signals were missed and for most cases no signals were missed; however, there was a small increase in the number of false positives.

We simulated ADRs occurring approximately within 1 month where the peak occurred at the middle of the month and we chose to perform the WSP test at the end of each month. When using the tool in practice, the appropriate cut points for the tool will depend on the duration of the study period. We also performed simulations using a lognormal and then uniform distribution for the times of events for ADRs. The resulting power of the WSP test when using these different distributions was very similar to those presented in the paper using the normal distribution. This indicates that when there are very few ADRs then the underlying distribution is inconsequential, and when there are a reasonable number of ADRs then the WSP test can detect these regardless of the underlying distribution of these events. We assumed that the standard deviation for the ADR distribution was the same regardless of the month it was centred on. It may be more reasonable to assume that the standard deviation would

Table II. Reliability of the Weibull Shape Parameter detection tool for a variety of cohort sizes, background and adverse drug reaction rates. Percentages calculated using 30 000 simulations

Cohort	Background event (%)	Adverse events, if present (%)[a]	Correct answers (%)	Incorrect signal generated [false positive] (%)	ADR missed [false negative] (%)	Specificity	Sensitivity
5 000	1	10	53	8	39	0.84	0.21
10 000	1	10	57	7	36	0.87	0.27
5 000	1	20	62	8	30	0.83	0.40
10 000	1	20	73	7	20	0.87	0.60
5 000	1	50	88	8	4	0.84	0.93
10 000	1	50	93	7	0	0.87	1.00
5 000	5	10	68	7	25	0.85	0.45
10 000	5	10	80	7	13	0.84	0.65
5 000	5	20	90	7	3	0.85	0.80
10 000	5	20	93	7	0	0.85	0.85
5 000	5	50	93	7	0	0.85	0.85
10 000	5	50	93	7	0	0.85	0.85
1 000	10	10	58	9	33	0.84	0.28
5 000	10	10	80	10	10	0.83	0.68
10 000	10	10	87	12	1	0.82	0.83
1 000	10	20	74	9	17	0.84	0.57
5 000	10	20	90	10	0	0.83	0.86
10 000	10	20	88	12	0	0.81	0.86
1 000	10	50	91	9	0	0.85	0.84
5 000	10	50	90	10	0	0.82	0.86
10 000	10	50	88	12	0	0.81	0.86

a ADRs are given as a percentage increase of the background rate of events.

ADR(s) = adverse drug reaction(s).

increase as duration from the start of the drug increases. The impact of this increased variation has not been assessed in this study.

The WSP test and WSP tool are not appropriate for the detection of rare ADRs (<1%) as the method requires that enough events occur in order to reliably estimate the hazard function. Spontaneous reports remain the most effective way to detect rare and serious ADRs.[10]

The simulations in our study did not include any loss to follow-up (i.e. censored observations for individuals) and this needs to be factored in when calculating the sample size for a study when using this method. Informative censoring will be an important issue when using time-to-event methods for analysing harm outcomes. The impact of this and potential adjustments to take this into account need to be investigated further.

Conclusions

The WSP test is a simple method to detect a non-constant hazard over time and can be used as a signal detection tool for cohort and EHR data. These preliminary simulations indicate that the test is most powerful at detecting ADRs that happen shortly after starting treatment. The power of the test can be increased by using a tool that repeats the WSP test at regular intervals. The WSP tool enables accurate detection of signals when the sample size is 5000 participants or more. The WSP test and tool is easy to implement using most standard statistical packages such as Stata and R. Further assessment of the test and tool on real rather than simulated data is required.

Acknowledgements

The authors declare no conflict of interest. No funding was received to perform this study.

References

1. Davis S, King B, Raine JM. Spontaneous reporting – UK. In: Mann R, Andrews E, editors. Pharmacovigilance. 2nd ed. New York: Wiley, 2007: 199-215
2. Goldman SA. Limitations and strengths of spontaneous reports data. Clin Ther 1998; 20 Suppl. C: C40-4
3. Bate A, Evans SJ. Quantitative signal detection using spontaneous ADR reporting. Pharmacoepidemiol Drug Saf 2009 Jun; 18 (6): 427-36
4. Shakir SAW. PEM in the UK. In: Mann R, Andrews E, editors. Pharmacovigilance. Chichester: John Wiley and Sons Ltd, 2007: 307-16
5. Coloma PM, Schuemie MJ, Trifiro G, et al. Combining electronic healthcare databases in Europe to allow for large-scale drug safety monitoring: the EU-ADR Project. Pharmacoepidemiol Drug Saf 2011 Jan; 20 (1): 1-11
6. Harmark L, Puijenbroek E, Grootheest K. Longitudinal monitoring of the safety of drugs by using a web-based system: the case of pregabalin. Pharmacoepidemiol Drug Saf 2011 Jun; 20 (6): 591-7
7. Hocine MN, Musonda P, Andrews NJ, et al. Sequential case series for pharmacovigilance. J Royal Statist Soc A 2009; 172: 213-36
8. Schuemie MJ. Methods for drug safety signal detection in longitudinal observational databases: LGPS and LEOPARD. Pharmacoepidemiol Drug Saf 2011 Mar; 20 (3): 292-9
9. Maignen F, Hauben M, Tsintis P. Modelling the time to onset of adverse reactions with parametric survival distributions: a potential approach to signal detection and evaluation. Drug Saf 2010 May 1; 33 (5): 417-34
10. Stephenson WP, Hauben M. Data mining for signals in spontaneous reporting databases: proceed with caution. Pharmacoepidemiol Drug Saf 2007 Apr; 16 (4): 359-65

Correspondence: Dr *Victoria Cornelius*, Senior Lecturer in Medical Statistics, Primary Care and Public Health Sciences, 7th Floor, Capital House, 42 Weston Street, London SE1 3QD, UK.
E-mail: victoria.cornelius@kcl.ac.uk

