DIA DEVELOP INNOVATE ADVANCE

Therapeutic Innovation & Regulatory Science
2016, Vol. 50(2) 188-194

DOI: 10.1177/2168479015602735
tirs.sagepub.com

Statistical Analysis of Cumulative Serious Adverse Event Data From Development Safety Update Reports

Brian Davis, MBChB[1], and Harry Southworth, PhD[2]

Abstract

Background: Monitoring safety in clinical trials by regulatory authorities and sponsors involves the clinical review, often subjectively, of large data sets of different types of safety information which may require considerable resources. *Methods:* This study investigated a means of statistically guided clinical review of safety data provided in the Cumulative Table of Serious Adverse Events of a Development Safety Update Report (DSUR). A simple statistical approach that treats every adverse event as independent of all others and uses a reference prior, which avoids infinite estimates of relative risk but does not unduly influence posterior inferences, was used with fixed rules of relative risk to identify a serious adverse event preferred term as a potential risk. *Results:* This simple model, using cumulative serious adverse event (SAE) data from 5 DSURs, identified a small group of potential risks that included some not reported by the sponsor as well as most of those reported by the sponsor in the DSUR Summary of Important Risks. *Conclusions:* The method provides a systematic and objective approach to analysis of cumulative SAE data that could help to identify potential risks that need further investigation by a regulatory authority or sponsor.

Keywords

clinical trials, signal detection, Bayesian hierarchical models, drug safety, regulatory authority

Introduction

The Development Safety Update Report (DSUR)[1] is an annual review and evaluation of safety information submitted to regulatory authorities to assure them that sponsors are adequately monitoring the safety profile of an investigational drug. It reports safety information obtained by the sponsor during the current review period and analyses the new information based on previous knowledge of the product's safety. The data for this study were taken from the Cumulative Table of Serious Adverse Events (DSUR Appendix VI)[2]; it contains cumulative serious adverse events (SAEs) reported to the sponsor, starting from the first authorization of a clinical trial with the investigational drug in any country up to the data-lock point for the DSUR.

An SAE is an event that results in death, is life threatening, results in hospitalization or prolongation of hospitalization, or causes persistent incapacity or congenital deformity.[3] They are usually reported as preferred terms (PTs) within system organ classes (SOCs), using the MedDRA dictionary,[4] which currently lists 26 SOCs, 334 high-level group terms (HLGTs), 1700 high-level terms (HLTs), and 20,559 PTs. It is common for several hundred PTs to be reported during any phase 2 or 3 clinical trial. Experience suggests that some PTs are effectively used as synonyms for each other; for example, patients reporting ''rash'' will seldom report ''dermatitis,'' and vice versa.

The information contained in the Cumulative Table of Serious Adverse Events in Appendix VI of a DSUR contains only a fragment of the available safety information on the investigational drug and contains data from pooling of results from different clinical trials, potentially with different indications, dosage forms, comparators, background therapies, and disease populations. As such, the information in Appendix VI is of quite poor quality, and because the individual studies are not identified, there is scope for Simpson's paradox to occur, where an unaccounted covariate causes an apparent treatment effect reversal. While moves are afoot to make more detailed clinical trial data available (EMA policy on publication of clinical data for medicinal products for human use,[5] for example), presently Appendix VI seems to provide the best available objective data on adverse events (AEs) for drugs in clinical development. The SAEs in Appendix VI are PTs listed alphabetically within

[1] Medicines and Healthcare Products Regulatory Agency (MHRA), London, United Kingdom
[2] Data Clarity Consulting Ltd, Stockport, United Kingdom

Submitted 26-Jan-2015; accepted 04-Aug-2015

Corresponding Author:
Brian Davis, MBChB, Fieldhead, Troutbeck, Penrith, Cumbria, CA110SH, United Kingdom.
Email: drbriandavis@btinternet.com

SOCs. This organization is not ideally suited to clinical review, as Appendix VI often includes hundreds of PTs presented on several pages. Further classification of PTs within HLTs within HLGTs and SOCs could help reviewers.

The development of methods for the analysis of AEs from clinical trials has challenged statisticians. Crowe et al[6] introduced the concept of a 3-tier approach, tier 1 being prespecified events, tier 2 common but not prespecified events, and tier 3 uncommon and not prespecified events. Mehrotra and Heyse[7] and Mehrotra and Adewale[8] used the false-discovery rate to alleviate concerns over multiple testing. Using the same data, Berry and Berry[9] developed a hierarchical model and compared results with *P* values from the Fisher exact test: further details are given below in the Methods section. Xia et al[10] reworked the Berry-Berry model using a Poisson likelihood rather than binomial, but Appendix VI of the DSUR does not contain detailed exposure information, so we do not consider it further. Southworth and O'Connell[11] reported a slightly simplified version of the Berry-Berry model that uses the same prior variance in each SOC and which is used in what follows. Southworth and O'Connell[11] also evaluated a number of measures for the association of an AE to treatment with the investigational drug. They showed that a penalized logistic regression method and an "inside-out" machine learning method worked well using AE data from a single clinical trial, but only the Berry-Berry approach is designed for aggregated data, so the other approaches cannot be applied to DSUR Appendix VI.

Part of the difficulty with taking a purely statistical approach to the analysis of AE data is that context, experience, data from other sources, and knowledge of plausible biological mechanisms must be taken into account. For example, known relationships between AEs and drugs in the same class or with similar biological pathways are clearly relevant when attempting to determine whether or not a potential relationship between a new investigational compound and a particular AE might be real. Also, some events, such as liver failure, are known to be associated with pharmaceuticals and a single case would suggest cause for concern despite the absence of "statistical significance." While it might be argued that such information could be built into a prior distribution of a Bayesian statistical model, to do so for 20,000 potential PTs, or for only the several hundred actually observed, would be a formidable task. Furthermore, synonymous PTs might only be identified after they have been observed, and even then it would likely require considerable subjective judgment as to whether they should be pooled. Statistical inference is only one part of the much larger process of scientific inference, and probability statements—either *P* values or Bayesian posterior probabilities—do not take all of the relevant context into account. As such, any probability statements from a statistical model can only be interpreted loosely. For these reasons, we consider statistical estimation of false discovery rates, and statistical multiplicity adjustments, to be of limited value: judgments as to which SAEs are related to treatment are not made on purely statistical grounds.

Methods

Given the difficulties in taking a purely statistical approach to the identification of SAEs caused by treatment in DSUR Appendix VI, we take on the modest task of ordering the SAEs using statistical analysis in such a way as to draw the attention of clinical reviewers to those that are likely to be of interest. Once such events have been identified, some will likely be dismissed, and some will require further investigation. In practice, such a ranking of SAEs ought to reduce potentially many pages of DSUR Appendix VI down to a number more easily manageable for clinical review.

The modest objective of this approach does not imply that statistical analysis has little role in the interpretation of safety data: for example, if liver failure was observed as an SAE, detailed statistical analysis of liver-related laboratory data would be appropriate.[12] However, such additional analyses require detailed patient-level data, and here we tackle only the cumulative DSUR Appendix VI data.

Statistical Methods

The Berry-Berry model for AEs seeks to take advantage of the hierarchical classification of PTs within SOCs. In particular, all PTs within a specific SOC are allowed to share information in such a way that the estimated event rates are drawn toward each other, given any apparent effect of treatment.

Berry and Berry[9] found their model to be sensitive to the allocation of AEs to body systems, and that it could be quite conservative in its estimation of treatment effects. In fact, because there are multiple competing hierarchies (eg, MedDRA Structured Medical Queries,[13] secondary SOCs, drug-specific hierarchies as determined by clinical trial sponsors or other interested parties), it is not necessarily the case that nesting PTs within SOCs provides the correct hierarchy.

Further, the Berry-Berry approach treats PTs within an SOC as being exchangeable. That is, knowing what treatment a patient received, it is assumed that no further information is available and each PT within an SOC is as likely to occur as any other PT within the same SOC. Clearly there are circumstances in which this assumption will not be true, but to provide information on the often hundreds of PTs reported in any particular trial, or the over 20,000 that exist in MedDRA, is impractical. A potential consequence is that the estimated effect of treatment on a PT or small group of related PTs is diluted by the lack of a treatment effect on many unrelated PTs in the same SOC. As DuMouchel[14] pointed out, in order to reasonably treat the PTs in this way, "the different safety issues should be medically related, so that it is plausible that the different issues have related mechanisms of causation or are different expressions of a broad syndrome." Besides PTs within SOCs being treated as exchangeable, the SOCs themselves are also treated as exchangeable, and the same concerns apply.

Bayesian approaches to statistical analysis require that prior distributions be provided for the quantities of interest. Berry and Berry[9] used prior distributions that have a spike at the point

where there is no treatment effect, mixed with a wide distribution that conveys little information. For any PT that is considered to be of interest a priori, the prior distribution used by Berry and Berry[9] will be inappropriate and will give undue weight to there being no treatment effect. Again, however, it will usually be impractical to go through all possible PTs in turn to specify appropriate prior distributions.

Evans et al[15] reported that the spiked priors in the Berry-Berry model result in complicated posterior distributions from which inference is not straightforward and proposed that vague prior distributions be used instead. Given all of these considerations, it was considered appropriate to compare the estimates from the Berry-Berry model with estimates from a much simpler approach.

We chose to estimate relative risk for each PT, although it can reasonably be argued that risk difference, odds ratios, or some other measure is more appropriate. In any event, it will always be important to look at the individual point estimates of risk. For the Berry-Berry model, the posterior predictive distribution can be used to compute point and interval estimates of relative risk.

Supposing that x events are reported from N_x patients in the experimental treatment group, that y events are reported from N_y patients in the comparator group, and that $\hat{p}_x$ and $\hat{p}_y$ are the estimated proportions of patients reporting the events, relative risk is defined to be

$$\hat{RR} = \hat{p}_x/\hat{p}_y.$$

The obvious estimate of $\hat{p}_x$ is x/N_x, and similarly for $\hat{p}_y$. However, when the number of events is zero in one of the treatment groups, this can lead to undesirable properties such as infinite or zero relative risks. In such cases, we can estimate $\hat{p}_x = (x + ½)/(N_x + 1)$ and similarly for $\hat{p}_y$. This approach is equivalent to adding 1/2 to all cells in the 2×2 table of event counts by treatment group. The method is suggested by Higgins and Green (Section 16.9.2).[16] For the case of a single proportion to be estimated, the approach is equivalent to using the Jeffreys prior and was studied by Brown et al[17] and found to have desirable properties. It can also be shown that this estimate of a proportion is equivalent to the predicted probability from a simple logistic regression model using Firth's bias-reduced procedure,[18] which is also the modified profile likelihood estimator (see, eg, Pawitan, Section 10.6).[19] The consequence of adding 1/2 to the numerator and 1 to the denominator when estimating proportions is that relative risk gets dragged closer to 1. For example, if there is 1 event out of 100 patients in the comparator group, and 2 of 100 in the experimental treatment group, the usual estimates of proportions lead to an estimated relative risk of 2. The adjusted estimates lead to an estimated relative risk of 1.67.

When there are zero events in the comparator group, the effect of using the adjusted estimates is more extreme because the relative risk using the unadjusted estimates is infinite and the adjusted relative risk gets dragged a great deal closer to 1. While this shrinkage of the relative risk toward 1 might be unintuitive, it will usually be unreasonable to conclude that a true relative risk is infinite, and the real story is that with rare events, it is difficult to estimate relative risk and the absolute risk is clearly low.

Given that this approach can be given a Bayesian interpretation, and that the Berry-Berry approach is explicitly Bayesian, we will refer to interval estimates as "credible intervals" rather than "confidence intervals." The Berry-Berry model works on the scale of the log odds ratio. However, the posterior predictive distribution of relative risk can be calculated from the Markov chains that result from the estimation procedure.

Following Sterne and Davey-Smith[20] and Cox,[21] we prefer to present 90% intervals rather than 95% intervals in order to prevent their abuse as surrogates for hypothesis tests at the 5% level. While a 90% interval can be used as a surrogate for a test at the 10% level, few researchers would consider this to be a convincing level of evidence, although experience suggests that many still consider 5% to be so. Following Cleveland (Section 3.14),[22] when presenting results graphically, we also provide the 50% interval estimates to give a better feel for the behavior of the central part of the distribution of relative risk (see the online supplementary material for example graph).

Assessment Method

The results of the simple model's calculations from the Cumulative Table of Serious Adverse Events (DSUR Appendix VI) were compared to the Sponsor's Summary of Important Risks (DSUR Section 19). Five DSURs were selected randomly by a scientific officer not connected with the study from all the DSURs received by the Medicines and Healthcare Products Regulatory Agency (MHRA) up to the end of 2011. To blind the statistical analyses, no information was provided to the statistician about the active drug, comparators, or their modes of action.

DSUR Appendix VI contains SAEs reported to the sponsor, starting from the first authorization of a clinical trial with the investigational drug in any country up to the data-lock point for the DSUR. They are presented either as a single table showing all SAEs from all trials with all forms of the drug or as individual tables of cumulative SAEs from each clinical trial. The columns of the table list the PTs grouped by SOC using the MedDRA classification system and the associated cumulative number of SAEs for the investigational drug and comparator. The data were anonymized and transposed into Excel worksheets (Microsoft), with the number of subjects exposed to the active drug and comparator at the top of the relevant column (see example data set in the online supplementary material).

The simple model calculated the relative risk for each SAE PT. The values of the point estimate of relative risk and the 90% and 50% credible bounds were presented as tables for (a) the 30 SAE PTs with the highest point estimate and (b) the 30 SAE PTs with the highest lower 90% credible bound (see example tables in the online supplementary material). An SAE

PT was considered a potential risk if either the point estimate of relative risk was $\geq$3.0, or its lower 90% credible bound was $\geq$1.0. All the potential risks identified by the statistical models were purely indicative and would need to be supported by more detailed analysis of the available data.

Potential risks selected by the simple model were compared to the sponsor's Summary of Important Risks in Section 19 of the DSUR, which can be provided as a table or text but does not include supporting data. Since the sponsor sometimes used an HLT rather than a PT to describe an identified or potential risk, the comparison was not precise; a sponsor's HLT could be related to several PTs selected by the simple model. To allow for this, the assessment considered that a PT had identified an HLT when any selected PT was related to a sponsor's HLT, and conversely that a sponsor's HLT had identified any related PT from the simple model. The comparisons for the 5 DSURs were summarized as the number and clinical nature of SAE PTs selected by the simple model that were and were not reported by the sponsor, and conversely as the sponsor's potential risks which were and were not selected by the simple model.

Results

The results from the Berry-Berry model when applied to the DSUR Appendix VI data quickly indicated a severe shortcoming. Many of the point estimates of relative risk, and one or both of the 90% credible bounds, were precisely 1. This feature is a symptom of the prior distributions with point mass at the log-odds of 0. Such posterior inferences are unintuitive and undesirable. Even when dealing with rare events about which there is considerable uncertainty as to any relationship with treatment, we would not wish to allow the prior distribution to dominate. For this reason, the Berry-Berry model was abandoned, and what follows is based only on the approach by the simple model that treats all PTs as being independent of each other.

The sources of the data for the 5 DSURs used in the analyses by the simple model are summarized in Table 1. It shows the number of clinical trials from which the cumulative SAE data were collected, the reported number of subjects exposed to the active drug and comparator, and the number of unique SAE PTs reported. In DSURs 3 and 5, because the sponsor reported cumulative SAEs for individual trials, a single clinical trial was selected for analysis by the simple model.

The comparison of the potential risks selected by the simple model and those reported by the sponsor is summarized in Table 2. It shows for the 5 DSURs: (1) the range of relative risk for the 30 SAEs with the highest level of relative risk, (2) the numbers of potential risks selected by the simple model as well as those reported and not reported by the sponsor, and (3) the numbers of potential risks reported by the sponsor and those selected and not selected by the simple model. In all the DSURs, the sponsor used HLTs to describe some potential risks and the simple model only used SAE PTs, the rules described above were used to compare the number of potential risks identified or not identified.

Table 1. Data sources used in the analyses by the simple model.

	Development Safety Update Report				
	1	2	3	4	5
Clinical trials, n	18	18	1	6	1
Subjects exposed, n					
Active	2310	3885	98	420	640
Comparator	1185	65	96	319	630
SAE PTs reported, n	349	729	36	302	308

PTs, preferred terms; SAE serious adverse event.

Some potential risks were identified by the sponsor from preclinical studies and postmarketing safety reports that were less likely to be detected by the simple model. Risks identified from these sources have been noted in the following summary of findings for each DSUR, as they might have affected the comparisons.

In DSUR 1, the investigational drug was not marketed. The sponsor reported 6 potential risks as 4 PTs and 2 HLTs. Two were based on preclinical data only. The simple model did not select the 2 risks based on preclinical data but selected 3 of the 4 risks based on clinical data; the one not selected was a nervous system disorder. The simple model selected 24 potential risks, from which the sponsor reported 4 but did not report the other 20. The potential risks not identified by the sponsor included renal failure, respiratory failure, and cardiac failure.

In DSUR 2, the investigational drug had been marketed in many countries for up to 8 years. The sponsor reported 9 identified or potential risks as 5 PTs and 4 HLTs. Three were based on preclinical data only. None was selected as a potential risk by the simple model; they were related to renal disease, hypersensitivity, and blood disorders. The simple model selected 2 potential risks, pyrexia and sepsis; neither was reported as a potential risk by the sponsor.

In DSUR 3, the investigational drug was not marketed. The data were taken from a single dose-escalation study. The sponsor reported 4 potential risks based on preclinical data as 3 PTs and 1 HLT. The simple model selected 3 but did not select 1 related to the cardiovascular system. The simple model selected 6 potential risks, of which the sponsor identified 2 but did not identify 4, including restless legs syndrome and hyperhidrosis.

In DSUR 4, the investigational drug was not marketed. The sponsor reported 8 risks as 5 PTs and 3 HLTs. The simple model identified 4 but did not identify the other 4. The risks not identified included bowel perforation, immune-mediated thrombocytopenia, and hypersensitivity reactions. The simple model selected all 30 SAE PTs as potential risks, of which the sponsor identified risks related to 12 but did not identify 18, including cardiorespiratory arrest, atrial fibrillation, renal failure, and pancreatitis.

In DSUR 5, the investigational drug had been marketed in 6 countries. The data were taken from a single clinical trial. The sponsor reported 22 potential risks as 18 PTs and 4 HLTs.

Table 2. Comparison of numbers of potential risks identified by the simple model and by the sponsor.

	Development Safety Update Report				
	1	2	3	4	5
Range of relative risk for 30 SAEs with highest values	7.70-2.57	4.64-0.53	6.86-0.54	64.61-5.32	25.69-2.85
Potential risks, n					
Selected by the simple model	24	2	6	30	29
Reported by the sponsor	5	0	2	12	16
Not reported by the sponsor	19	2	4	18	13
Reported by the sponsor	6	11	4	7	22
Selected by the simple model	3	0	3	4	13
Not selected by the simple model	1	11	1	3	9

SAE, serious adverse event.

Three were identified only from postmarketing data. The simple model selected 13 but did not select 9. The risks not identified included pulmonary embolism, thrombocytopenic purpura, and increase of liver enzymes. The simple model selected 29 potential risks, of which the sponsor identified risks related to 16 but did not identify 13. The risks not identified included peripheral and central nervous disease, sudden death, and hyperkalemia.

Discussion

As discussed in the introduction, the Cumulative Table of Serious Adverse Events in a DSUR has various limitations. Therefore, any statistical analysis using only cumulative SAE data must be treated as purely exploratory and could not, by itself, be used to draw strong conclusions. However, the results suggest that the method could help to select a small number of SAE PTs for further assessment as potential risks. As such, the approach has potential to objectively guide clinical review of DSUR data.

A level of relative risk was chosen that was thought to provide a plausible indicator of a potential risk that needed further consideration. An SAE was considered a potential risk if either the point estimate of relative risk was $\geq$3.0 or its lower 90% credible bound was $\geq$1.0. However, the results showed that only 1 of the SAEs had a lower 90% credible bound $\geq$1.0 without a point estimate $\geq$3.0. Experience should lead to refinement of the rules and possibly different rules for different indications. Indeed, the experience with DSUR 4 illustrates that in some cases it will be necessary to look at more than 30 PTs, a number that was chosen in advance on the grounds that experience suggested it would be high enough, but which turned out to be too low in that example.

As discussed previously, the data that appear in DSUR Appendix VI is a mixture from more than 1 trial, and so all analyses need to be interpreted with caution. Relative risk was used in this study as a convenient summary statistic to be interpreted in the context of other available information. Its use does not imply that it will always be the only appropriate estimate of a treatment difference. In some cases, the absolute difference in risk might be more sensible, and in all cases the individual point estimates of risk will be relevant. The estimate of relative risk by the simple model can only identify potential risks and needs to be supported by more detailed analysis of the available data. Nevertheless it does provide an independent objective approach to analyzing cumulative SAEs reported in a DSUR.

The potential risks identified by the simple model were compared to the sponsor's identified and potential risks as a means of comparing the simple statistical method with the current method of safety assessment. In all the DSURs, the sponsor used HLTs rather than PTs to identify some risks, which made the comparison of the sponsor's HLT risks to the simple model's PT risks somewhat subjective. This limitation could be avoided if sponsors only used PTs to describe identified and potential risks. The simple model and rules selected most of the potential risks identified by the sponsor in DSURs 1, 3, 4, and 5, but not in DSUR 2, in which the simple model did not select any of the 9 risks identified by the sponsor. This discrepancy might have occurred because the sponsor had additional safety information from the drug, which had been marketed for many years. Also, some sponsors' potential risks were based on preclinical studies and might not be detected from SAE data. In DSURs 1, 2, and 3, some potential risks were based only on preclinical data, and interestingly, in DSUR 3, the simple model selected 3 of 4 of these but none in the other DSURs.

The study used cumulative SAE data from 5 first DSURs for investigational drugs. Studies in the future might compare the analyses of DSURs from one year to the next to see if they can provide additional objective approaches to the analysis of the safety of an investigational drug.

This study was designed to assist assessors in a regulatory authority to objectively assess the large number of DSURs received; for example, the MHRA receives approximately 100 DSURs per month. To help with this, the analysis could be automated from a standardized spreadsheet of the Cumulative Table of Serious Adverse Events, or a sponsor could be asked to submit tables listing the cumulative SAE PTs with estimated relative risk and lower 90% credible bounds above certain values. The data could also be presented as graphs to facilitate assessment.

Conclusion

The Berry-Berry model produces unintuitive posterior inferences as a consequence of the choice of prior distribution. While it could be argued that using a more traditional vague Gaussian prior would be appropriate, the exchangeability assumption among PTs within each SOC, and among the SOCs, is highly questionable. While it is also questionable to suppose that all PTs are independent of each other, in practice it is difficult to specify in advance which should be treated as exchangeable and which should not.[14] Once the clinical reviewer has identified PTs requiring further investigation, it might be appropriate, albeit controversial, to then identify relevant groupings and perform more sophisticated analyses. Furthermore, it is hoped that the approach outlined here will be implemented by regulators tasked with reviewing DSURs, and the computational complexity of a 3-level hierarchical Bayesian model is a barrier to adoption.

The simple model provides an objective approach to analysis of cumulative SAE data. The results suggest that relative risk can be used to focus on a small group of SAEs as potential risks. They identified a number of SAEs not reported by the sponsor that would need further assessment and in most cases selected the SAEs reported as potential risks by the sponsor. This objective approach could help an assessor in a regulatory authority to decide whether a sponsor was adequately monitoring and evaluating the evolving safety profile of an investigational drug and a sponsor to identify potential risks that warrant further investigation.

Author Note

This article represents the views of the authors and should not be construed to represent the views of the MHRA.

Acknowledgments

The authors thank MHRA staff Ian Hudson, Martyn Ward, and Robert Hemmings for helping to initiate this study, Balall Naeem and Sophia Hussain for assistance with the data, and the referees for helpful suggestions that materially improved the manuscript.

Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: B. Davis was employed by MHRA for 20 years and has no other affiliations. He was responsible for maintaining confidentiality of the data for this study; no information was provided about the active drug, comparators, and their modes of action for the statistical analysis. H. Southworth is an independent statistical consultant and performs work for various pharmaceutical companies.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: B. Davis was funded by an MHRA Research Contract for this study. H. Southworth was employed by AstraZeneca during the course of most of the work.

Supplemental Material

Online supplemental material for this article is available on the journal's website at http://tirs.sagepub.com/supplemental.

References

1. ICH Guideline E2F, Development safety update report. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E2F/Step4/E2F_Step_4.pdf. Published 2010. Accessed June 16, 2015.
2. ICH Guideline E2F, Example development safety update report. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E2F/Examples_DSUR/E2F_Example_Commercial_DSUR.pdf. Published 2010. Accessed June 16, 2015.
3. ICH Guideline E2A, Clinical safety data management: definitions and standards for expedited reporting. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E2A/Step4/E2A_Guideline.pdf. Published 1994. Accessed June 16, 2015.
4. MedDRA Hierarchy. http://www.meddra.org/how-to-use/basics/hierarchy. Accessed June 16, 2015.
5. European Medicines Agency policy on publication of clinical data for medicinal products for human use. EMA/240810/2013. http://www.ema.europa.eu/docs/en_GB/document_library/Other/2014/10/WC500174796.pdf. Published October 2, 2014. Accessed June 16, 2015.
6. Crowe B, Xia HA, Berlin J, et al. Recommendations for safety planning, data collection, evaluation and reporting during drug, biologic and vaccine development: a report of the Safety Planning, Evaluation and Reporting Team (SPERT). *Clin Trials*. 2009;6:430-440.
7. Mehrotra DV, Heyse JF. Use of false discovery rate for evaluating clinical safety data. *Stat Methods Med Res*. 2004;13:227-238.
8. Mehrotra DV, Adewale AJ. Flagging clinical adverse experiences: reducing false discoveries without materially compromising power for detecting true signals. *Stat Med*. 2012;31:1918-1930.
9. Berry SM, Berry DA. Accounting for multiplicities in assessing drug safety: a three-level hierarchical mixed model. *Biometrics*. 2004;60:418-426.
10. Xia HA, Ma H, Carlin BP. Bayesian hierarchical modeling for detecting safety signals in clinical trials. *J Biopharm Stat*. 2011;21:1006-1029.
11. Southworth H, O'Connell M. Data mining and statistically guided clinical review of adverse event data in clinical trials. *J Biopharm Stat*. 2009;19:803-817.
12. Southworth H, Heffernan JE. Extreme value modelling of laboratory safety data from clinical studies. *Pharma Stat*. 2012;11:361-366.
13. Introductory Guide for Standardised MedDRA Queries (SMQs). Version 16.0. http://www.meddra.org/sites/default/files/guidance/file/smq_intguide_16_0_english.pdf. Accessed June 16, 2015.
14. DuMouchel W. Multivariate Bayesian logistic regression for analysis of clinical study safety issues. *Stat Sci*. 2012;27:319-339.
15. Evans SJW, Prieto-Merino D, Spiegelhalter DJ, Whittaker J, Smeeth L. Implications of using different types of priors in a three-level hierarchical Bayesian model for the analysis of drug

adverse events, presented at the International Biometrical Conference, Dublin, July 13-18, 2008. http://ibc2008abstracts.tibs.org/Bayesian%20Methods/IBC2008_Paper_ref_429.pdf. Accessed June 16, 2015.
16. Higgins JPT, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0. The Cochrane Collaboration, 2011. www.cochrane-handbook.org.
17. Brown LD, Tony Cai T, DasGupta A. Confidence intervals for a binomial proportion and asymptotic expansions. *Stat Sci*. 2001; 16:101-133.
18. Firth D. Bias reduction of maximum likelihood estimates. *Biometrika*. 1999;80:27-38.
19. Pawitan Y. *In All Likelihood: Statistical Modelling and Inference Using Likelihood*. Oxford: Clarendon Press; 2001.
20. Sterne JAC, Davey Smith G. Sifting the evidence—what's wrong with significance tests? *Br Med J*. 2001;322:226-231.
21. Cox DR. Another comment on the role of statistical methods. *Br Med J*. 2001;322:231.
22. Cleveland WS. *The Elements of Graphing Data*. Summit, NJ: Hobart Press; 1994.