Statistical Science
2012, Vol. 27, No. 3, 346–347
DOI: 10.1214/12-STS381C
Main article DOI: 10.1214/11-STS381

An Answer to Multiple Problems with Analysis of Data on Harms?

Stephen Evans

Key words and phrases: Multiple outcomes, drug safety.

Randomized Controlled Trials have had many statistical developments since their introduction into modern medicine sixty years ago. While many of the advances have addressed general design and analysis issues, most have been motivated by or focused on assessment of efficacy in contrast to safety. The reporting of harms is demonstrably weak (Ioannidis and Lau, 2001; Loke and Derry, 2001) and, in spite of the CONSORT guideline on harms (Ioannidis et al., 2004), continues to show some deficiencies (Pitrou et al., 2009). Analytical developments focused on harms have been very limited, reflecting lack of statistical effort employed in that area as well as perhaps a general neglect of the less exciting area of safety.

DuMouchel made a major contribution to extracting useful information from spontaneous reports using Bayesian methods (DuMouchel, 1999; DuMouchel and Pregibon, 2001). This paper (DuMouchel, 2012) is potentially an important advance in assessing data on harms from randomized trials. As an incidental point, it is possible it will also have application in observational studies as well.

There are several really important features of MBLR as set out by DuMouchel:

(1) It addresses a clinically relevant problem, not addressed by standard methods. The problem being that it is difficult if not impossible to prespecify possible harms in terms of formal hypotheses and the multiple medically related issues need to be seen as a broad picture as well as being reflected by narrow medical terms. In practice, also the data may be very limited because serious harms are rare for medicines reaching the market.

(2) It addresses at least part of the problem of multiplicity of many possible hypotheses of harm.

Stephen Evans is Professor, Department of Medical Statistics, London School of Hygiene & Tropical Medicine, Keppel Street, London, WC1E 7HT, UK (e-mail: stephen.evans@lshtm.ac.uk).

(3) It avoids the epidemiological dilemma of lumping or splitting terms which can lead to reduced sensitivity or simple loss of statistical power. It also avoids the difficulty caused by composite outcomes which, although they have their place in assessing efficacy, have problems in that context but potentially worse problems in the context of safety.

(4) It provides medically useful and interpretable estimates of effects while retaining a good statistical foundation. The modeling is consistent for each response variable related to a possible harm, and can be used in trials which are primarily aimed at testing for efficacy.

(5) It does not seem heavily reliant on the particular form of the Bayesian Priors being used.

The potential of the method is therefore very considerable and seems destined to be used by the pharmaceutical industry and may eventually be encouraged by regulators if it is shown in practice to be useful, applicable and reasonably easily implemented.

Nothing in life is perfect though! It does require prespecification of a group of medically-related terms, expressed as simple binary responses and expected to behave in a similar manner (on a relative or odds ratio scale)—showing exchangeability in Bayesian parlance. This may not always be simple to do in practice, and even with the use of a hierarchical medical dictionary like MedDRA which has over 16,000 "preferred terms," the choice of the number of terms and how wide a range is included will not always be easy. There is then a danger that a nonprespecified analysis may reach the conclusion desired by the analyst or sponsor. It will be interesting to see if the MBLR method can be applied to the more limited number of terms in the internationally agreed "Standardized MedDRA Queries" which are already groupings of terms from one or more MedDRA System Organ Classes (SOCs) that relate to a defined medical condition or area of interest. Similarly, it may be possible to use a form of cluster analysis to provide medically sensible groupings based on one set of data (not necessarily from RCTs) and then