Estimating Event Rate Differences Using Data from Blinded Trials: The Canary in the Coal Mine

A. Lawrence Gould[1], William W. B. Wang[1]
[1]Merck Research Laboratories P.O. Box 1000, Upper Gwynedd PA 19454-2505

Abstract

The development of drugs and biologicals whose mechanisms of action may extend beyond their target indications has led to a need to identify unexpected potential toxicities even in ongoing blinded clinical trials. Recently issued FDA rules regarding safety reporting requirements raise the possibility of breaking the blind for serious adverse events that are not the clinical endpoints of a blinded trial. However, unblinding individual cases of frequently occurring adverse events could compromise the overall validity of the trial. The possibility of elevated risk can be addressed without unblinding the trial by using external information about adverse event rates among patients not receiving the test product in populations similar to the trial population. We describe a Bayesian approach to determining the likelihood of elevated risk suitable for binomial or Poisson likelihoods that applies regardless of the metric used to express the difference. The method appears to be particularly appropriate when the adverse events are not 'rare'.

Key Words: Bayesian, parametric, *p*-value, Fisher

1. Introduction

Assurance of the safety of drugs, vaccines, and other medical products has been an explicit regulatory objective for more than a century. With the development of drugs and biologicals whose mechanisms of action may extend beyond their target indications, there has been an increasing recognition of the need to detect potential toxicities as soon as possible. This need extends to identifying unexpected potential toxicities even while blinded clinical trials are under way during the product development process.

The FDA recently issued rules regarding safety reporting requirements pertaining to products still under development. [1] One component of these rules is the possibility of breaking the blind for serious adverse events that are not the clinical endpoints of a blinded study in order to identify a possible risk associated with the new product. A number of concerns were raised in comments responding to the rule, including the possibility that unblinding individual cases of frequently occurring adverse events could compromise the overall blinding of the study.

It may be possible to obtain information about relative adverse event risks of a new product without unblinding individual cases, especially if the trials are large. If a trial is blinded, then only the total number of patients in the trial and the total number of reports of an adverse event are known. The actual numbers of patients on the test and control treatments are unknown, as are the numbers of patients on each treatment who report the adverse event.

The key question is, for which adverse events should consideration be given to taking some action such as partially unblinding individual cases to conform to the recently

issued rules? Typically, these would be the adverse events that appear to be occurring unusually frequently. However, 'unusually frequently' has to be considered in context, particularly with regard to the patient population included in the trial.

Literature databases contain few articles dealing with estimates based on blinded data. Blinded trials do not provide reliable estimates of differences between treatment effects [2, 3], so using only blinded data provided by an individual trial or suite of trials is unlikely to provide useful insights into potential toxicity issues.

External information about the natural incidence of an adverse event in the clinical trial population, and some simplifying assumptions, can be used to provide treatment risk difference estimates that are insufficiently accurate to justify firm conclusions, yet may be useful for identifying potential toxicity issues for which further followup would be appropriate. An intuitive approach in this direction is provided in [4].

2. Model

2.1 Single Trials

The information provided from a typical trial, or from a collection of trials, consists of the allocation fraction, the numbers of patients assigned to the test and control treatments, the numbers of events reported by the patients on each treatment, and the test and control event rates (p_T, p_C, respectively). If the trial is blinded, then the only known quantities are N, the total number of patients in each trial, X, the total number of reports of the event by the patients in either treatment group of each trial, and θ, the fraction of the total sample size allocated to the test group.

The description that follows applies to a single trial, but can be extended to multiple trials. If the event counts are generated from binomial likelihoods, then the joint likelihood is

$$\text{Likelihood} = f_{\text{binom}}(X;\, N,\, p_C)\left(\frac{1-p_T}{1-p_C}\right)^{\theta N} C(x_T, X, N, \theta)\left(\frac{p_T(1-p_C)}{(1-p_T)p_C}\right)^{x_T} \qquad (1)$$

where $C(x_T, X, N, \theta)$ denotes a combinatorial factor. If M denotes a measure of the difference between the test and control event rates, e.g., the ratio $R = p_T/p_C$, then p_T in (1) can be replaced by a function of M and p_C. Summing (1) with respect to x_T gives the probability of X as a function of p_C and M. Multiplying the result by the product of the prior densities for M and p_C and integrating with respect to M and p_C gives the marginal probability function of X that can be used to construct control charts for monitoring ongoing trials.

If N is large and p_C and p_T are small, then the events can be regarded as realizations from Poisson likelihoods and the analogue of (1) is, when $M = R = \lambda_T/\lambda_C$ is the ratio of event rates,

$$\text{Likelihood} = f_{\text{Poisson}}(X;\, (\phi R + 1)\lambda_C) \times f_{\text{binom}}(x_T;\, X,\, \phi R/(\phi R + 1)) \qquad (2)$$

where ϕ is the exposure of subjects on T relative to those on C. Exposure can be expressed in terms of the allocation ratio, $\phi = \theta/(1 - \theta)$, duration of treatment, etc.

The next step is to specify priors for M and for p_C, usually on the basis of literature reports and prior data reflecting regulatory and clinical considerations. The prior distribution for M should be diffuse, so that its posterior is driven primarily by the observed event count, but the prior distribution for p_C should be precise. Conventional Bayesian calculations generate the posterior density of M given the observed event count and the parameters of the prior distributions of M and p_C, from which one can calculate the posterior probability that the metric exceeds some critical value, e.g., the posterior

probability that the ratio $R = p_T/p_C$ exceeds 2. If this probability is sufficiently high, then some action such as partial unblinding might be considered.

2.2 Determining the Prior Distributions

Information about the prior distribution of the control group event rate p_C can be obtained from various sources, e.g., large electronic record health databases. This information can be used in various ways, typically involving regulatory and clinical input, to specify a prior distribution for p_C Figure 1 illustrates diffuse (dashed curve) and precise (solid curve) prior beta(a,b) distributions for p_C when the expected event rate is 2% or 10%.

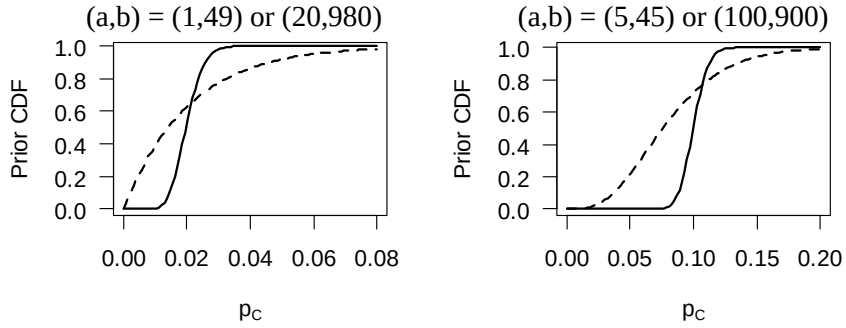

Figure 1: Prior cdfs for the control group event rate pC assuming an expected rate of 2% (left panel) or 10% (right panel)

Since, as will be demonstrated, the sensitivity of the method for detecting possible increases in risk depends strongly on the location and precision of the prior distribution of p_C, careful attention should be given to its determination. It generally will be sufficient to assume that a transformed value of M has a normal prior distribution with low precision.

2.3 Decision Rule

A typical decision rule might be "Consider taking action if $P_{post}(M > M_{crit} \mid X) > \gamma_{crit}$". The posterior probability depends on the total sample size N and the total number of events X. The values of M_{crit} and γ_{crit} need to be determined before observing the trial data. Table 1 displays the values of $X_{crit} = \min(X : P_{post}(M > M_{crit} \mid X) > \gamma_{crit})$ when M is the rate ratio $R = p_T/p_C$. Observing X_{crit} or more events implies the posterior probability will exceed γ_{crit}.

Table 1: Values of $X_{crit} = \min(X: P_{post}(R < R_{crit} \mid X_{crit}) > \gamma_{crit})$ for various prior distributions of p_C when $(R_{crit}, \gamma_{crit}) = (1, 0.8)$, (2, 0.5), or (2, 0.8).

	$E(p_C)$	0.02	0.02	0.1	0.1
$(R_{crit}, \gamma_{crit})$	Prior (a,b)	(1, 49)	(20, 980)	(5, 45)	(100, 900)
(1, 0.8)	N = 100	14	4	22	13
	1000	56	31	134	105
(2, 0.5)	100	18	8	28	19
	1000	136	38	241	156
(2, 0.8)	100	30	12	40	23
	1000	238	50	344	172

2.4 Statistical Properties

The statistical properties of the approach are expressed in terms of the probability that the observed total event count X will exceed a critical value X_{crit} such as in Table 1 as a

function of the true values of p_C and p_T. Figure 2 provides a graphical summary of the statistical properties of the method under various assumptions about the true values of p_C and R. The vertical lines identify the X_{crit} values corresponding to (R_{crit}, γ_{crit}) = (2, 0.5) for a prior with low precision ('diffuse') or with high precision ('sharp'). The ordinate corresponding to the intersection of a vertical line with a curve corresponding to a true value of R. which is unknown, gives the probability of observing an X value no less than X_{crit}, i.e., of possibly taking some action, for that true value of R.

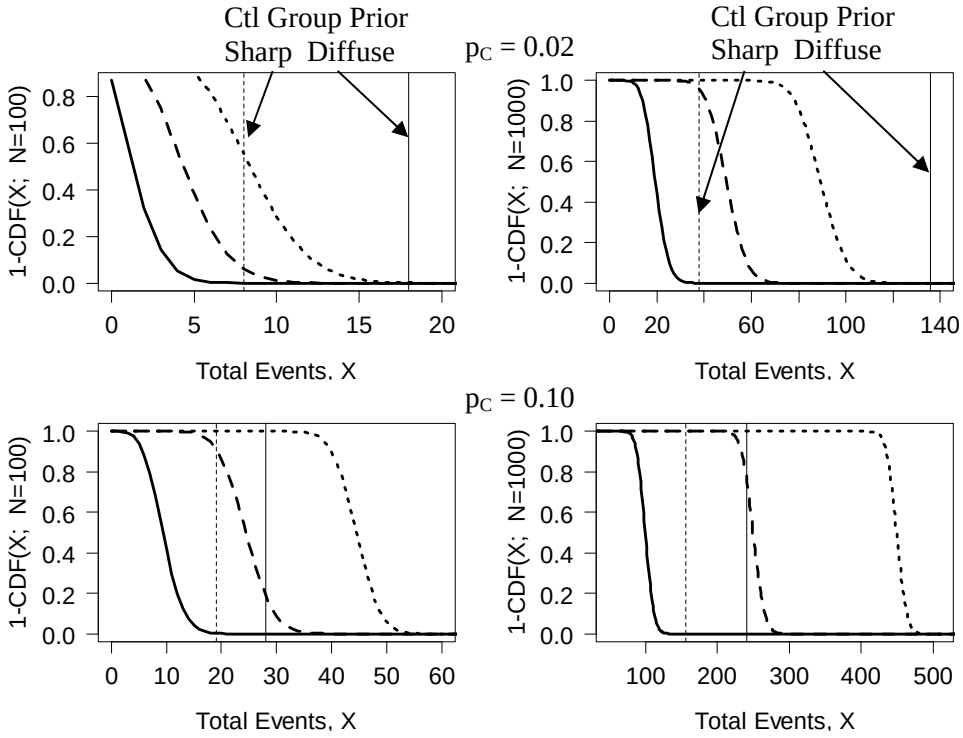

Figure 2: Upper tails of the probability distribution of X as a function of the total sample size N (left panel: N = 100, right panel: N = 1000), the expected control group event rate p_C (upper panels: p_C = 0.02, lower panels: p_C = 0.10), the spread of the prior distribution of p_C, and the true value of the event rate ratio R = p_T/p_C [solid: R = 1, dashed: R = 2, dotted: R = 4] The vertical lines correspond to the values of X_{crit} when (R_{crit}, γ_{crit}) = (2, 0.5).

It is evident from Figure 2 that the probability of an observed value of X exceeding X_{crit} depends on the total sample size, the precision of the prior distribution of the control group event rate, and the true value of R. If the prior distribution of the control group event rate is 'diffuse' and the expected value of p_C is small (0.02), then the probability that X exceeds X_{crit} is essentially zero regardless of the sample size even if R = 4. If the prior distribution of the control group event rate is precise, then the probability that X exceeds X_{crit} is nearly 1 for R = 2 or 4 except when N = 100 and the expected value of p_C is 0.02, when the probability is about 0.10. If R = 1 (solid curve), corresponding to no risk inflation, then there is essentially zero probability that X > X_{crit} for any of the cases.

3. Example

We illustrate the method in the context of trials in women with Type 2 diabetes mellitus, with the target adverse event being atrial fibrillation. Table 2 summarizes information about the 'natural' incidence of atrial fibrillation among these patients obtained from the

literature and from large databases. For the sake of illustration, the findings in Table 2 are assumed to provide a reasonable assessment of the range of values for the probability that a woman with Type 2 diabetes mellitus will develop a first occurrence of atrial fibrillation in a year of observation. The objective is to determine an appropriate sample size for a new trial that provides reasonable sensitivity for detecting potential toxicity of a treatment for Type 2 diabetes mellitus.

Table 2: Incidence of atrial fibrillation among women with Type 2 diabetes mellitus but without a previous history of atrial fibrillation.

Source	No. Pats	Incidence/1000 Patient-Years
CPRD	66,036	6.47
CCMC	2,500,051	13.95
Schoen *et al* [5, 5]	3,667	3.67
Nichols *et al* [6, 6]	7,836	7.6

Suppose that 10,000 patients are to be observed for a year in each treatment group of the new trial. The sample size has to be large because atrial fibrillation is relatively rare. Since the natural event rate (incidence per person-year) p_C is quite small, a Poisson model like (2) will be appropriate. The findings in Table 2 suggest (details immaterial) a beta distribution for p_C with parameters (75, 8366), which has about 95% of its probability content between 0.007 and 0.011. This corresponds to a Poisson event rate parameter $\lambda_C = 10{,}000p_C$ with a gamma (75.75, 0.85) prior distribution. This distribution has the central 95% of its probability between 70 (= 0.007 × 10,000) and 110 (= 0.011 × 10,000).

Figure 3 summarizes the statistical properties for this example assuming that the true value of λ_C is 70, 90, or 110. The vertical line is at the value of X_{crit} corresponding to $R_{crit} = 2$ and $\gamma_{crit} = 0.5$. A decision to take some action ($X > X_{crit}$) if the true event ratio was 2

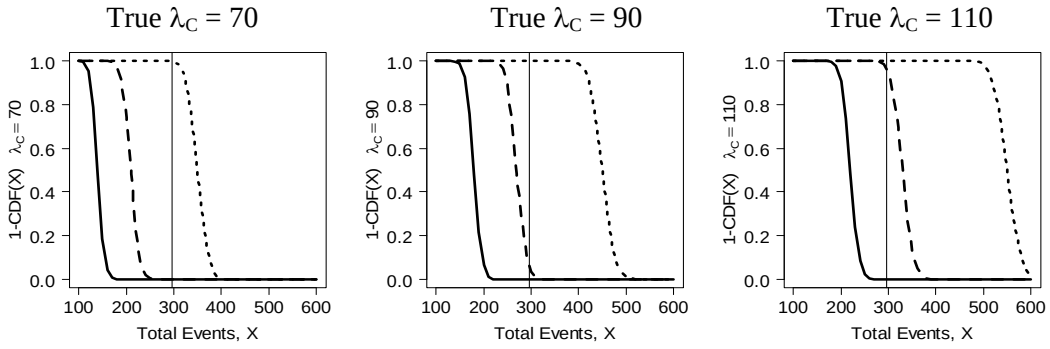

Figure 3: Distribution of the total event count X as a function of the control group event rate λ_C and the true ratio R of the test and control group event rates. The solid curve corresponds to R = 1, the dashed curve to R = 2, and the dotted curve to R = 4.

would be highly unlikely if the true value of λ_C is 70 or 90, but almost certain if the true value of λ_C is 110. It is clear from from Figure 3 that the prior distribution of the control group event rate should be precise, to minimize the prior probability of values of λ_C that would lead via the likelihood to an excess number of events in the control group and, consequently, a total number of events implying an elevated event risk in the test group for which some action should be contemplated even if the expected event rates were the same in both groups or the test group event risk was only modestly increased.

4. Discussion

The key conclusion is that it may not be necessary to unblind an ongoing trial to satisfy the spirit of the recently issued FDA rules regarding safety reporting requirements pertaining to products still under development [1] if external information about adverse event rates in relevant patient populations can be obtained from large databases such as electronic health databases. However, unless the prior information provides a very precise estimate of the control group event rate, the trial to which the method would be applied is fairly large, and the increase in risk from the test agent is substantial, a finding that would justify unblinding the trial is unlikely. This is not necessarily a disadvantage, because unblinding a trial has many consequences and it would be inadvisable to do so without substantial evidence suggesting that it was necessary.

The illustration of the statistical properties of the method provided in Section 2 is based on parameters (chosen for illustration) for the prior distributions that define the method. In practice, the values of the parameters rest on clinical and regulatory considerations and must reflect discussion and concurrence of development program sponsors and managers, and regulators. It is unlikely that a single set of parameter values will apply in general or even within a particular development program.

Various rules can be described for specifying actions that could be taken based on the application of the method described here. Regardless of the action that any rule might prescribe, the occurrence of larger numbers of adverse events than would be anticipated from prior experience cannot by itself be interpreted as demonstration of increased risk from the test treatment. Associative or causal treatment relationships can be demonstrated only when the trial is unblinded and the treatment assignments are known.

The method is intended to be used primarily for screening large ongoing blinded trials to identify situations where the incidence of adverse events is substantially larger than would be anticipated on the basis of prior information about the 'natural' incidence of the events, that is, the incidence for patients in the control group. Its purpose is to alert trial sponsors and managers to a potential safety issue in a principled way without compromising the blinding or integrity of the trial. In that sense, it is analogous to a control chart in a manufacturing process, or a canary in a coal mine, and so can be applied repeatedly during a trial. It is not intended to, and cannot, provide a precise or even necessarily accurate ongoing assessment of the relative test/control adverse event risk. Assessment of the reason for an elevated overall event count requires at least partial unblinding and a separate analysis.

References

1. FDA. *Guidance for Industry and Investigators: Safety Reporting Requirements for INDs and BA/BE Studies* Food and Drug Administration, 2012 http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM227351.pdf . Accessed 12-3-2014.

2. Miller F, Friede T, Kieser M. Blinded assessment of treatment effects utilizing information about the randomization block length. *Statistics in Medicine* 2009; **28**:1690-1706.

3. Zhou K, Ganju J. Inference from blinded data in randomized clinical trials. *Clinical Trials* 2013; **10**:744-753.

4. Ball G, Piller I, Silverman M. Continuous safety monitoring for randomized controlled clinical trials with blinded treatment information. *Contemporary Clinical Trials* 2011; **32**:S2-S17.

5. Schoen T, Pradhan AD, Albert CM, Conen D. Type 2 diabetes mellitus and risk of incident atrial fibrillation in women. *Journal of the American College of Cardiology* 2012; **60**:1421-1428.

6. Nichols GA, Reinier K, Chugh SS. Independent contribution of diabetes to increased prevalence and incidence of atrial fibrillation. *Diabetes Care* 2009; **32**:1851-1856.