Statistics and Its Interface Volume 11 (2018) 327–337

Does an observed zero-total-event study contain information for inference of odds ratio in meta-analysis?*

Min-ge Xie†, John Kolassa, Dungang Liu, Regina Liu, and Sifan Liu‡

This note is concerned with the contribution of an observed *zero-total-event study*, defined to be a study which observes zero events in both treatment and control arms, in meta-analysis. It provides a comparison of two approaches, namely the regular likelihood approach and the classical conditional likelihood approach, from several perspectives. This topic has long been debated, and it has received much renewed interest recently, in part due to the divergent views on the handling of zero-total-event studies in the high profile publication Nissen and Wolski (2007). Following a careful study of both approaches and an illustration of a numerical example, we find that, when we assume the underlying population event rates are not zero, an observed zero-total-event study actually contains information for inference on the parameters such as the common odds ratio in meta-analysis and cannot be left out in our analysis. This is contrary to the belief held by many statisticians that an observed zero-total-event study does not contribute to meta-analysis because it does not contain any information concerning the common odds ratio. The latter belief is mainly formed based on conditional likelihood arguments and/or that an observed zero-total-event study alone cannot provide a meaningful confidence interval for the odds ratio. Our finding should help clarify a difficult question concerning how to deal with zero-total-event studies in meta-analysis of rare event studies.

Keywords and phrases: Clinical trials, Conditional inference, Likelihood, Meta-analysis, Rare event, Two-by-two table, Zero-total-event study.

1. MOTIVATION AND INTRODUCTION

For a binomial experiment with an unknown event rate π as the parameter of interest, observing a zero event out of 1,000 trials provides a different inference about π than observing a zero event out of 10 trials. Although the point estimates of π are zero in both cases, the former case typically provides a lower upper confidence bound π (see, for example, Hald, 1952, Sections 14.4 and 21.3) and hence a stronger evidence showing that π is close to zero than that from the latter case. However, when data consist of a series of pairs of binomial variables arranged in two-by-two tables, and one desires inference on certain parameters comparing the event probabilities for the coupled binomial observations, the appropriate method for using information in tables with zero events is far from settled (see, e.g., Finkelstein and Levin, 2012).

In the analysis of two-by-two tables, a study is referred to as a *zero-total-event study* if zero events are observed in both the treatment and control experiments (c.f., Sweeting et al., 2004; Bradburn et al., 2007). The parameter of interest in a treatment-versus-control two-by-two table is often its associated risk difference $\pi_1 - \pi_0$, risk ratio π_1/π_0 or odds ratio $\{\pi_1/(1-\pi_1\}/\{\pi_0/(1-\pi_0)\}$. Here, π_0 and π_1 are the underlying population event rates in the treatment and control group, respectively. As in the single binomial case, there is some general agreement that the zero-total-event studies with different sample sizes provide different information about the risk difference (measured by confidence bounds, for instance), although the point estimates of the risk difference are all zero. Tian et al. (2009) provided a nice exact inference procedure to harvest information from zero-total-event studies to make inference for risk difference in a meta-analysis. However, for other parameters such as risk ratio and odds ratio, there are divergent views on what the correct inference statement should be when zero-total-event studies are observed (cf., Finkelstein and Levin, 2012 and reference therein). In meta-analysis, in particular, the presence of zero-total-event studies has long been considered a challenge, and how or whether these studies can be effectively incorporated into meta-analysis has been hotly debated (see, e.g., Cai et al., 2010; Finkelstein and Levin, 2012). This debate has become even more heated in recent years, in part fueled by the divergent views over the handling of rare event studies in the high profile publication Nissen and Wolski (2007). Many statisticians, scientists and policy makers have all been actively investigating whether or not zero-total-event studies should be included in meta-analysis. Some favor exclusion, claiming that a zero-total-event study does not contain any information about the parameter common

*We thank the editor and reviewers for their constructive comments, which helped improve the manuscript.
†Corresponding author.
‡The authors acknowledge support from grants: NSF-DMS 1513483, 1107012; and NSA:11-1-0157;14-1-0102. The third author also acknowledges the support from NIH (R01 DA016750-09).

odds ratio. This policy is supported by the conditional likelihood formulation. Others argue for inclusion, even though there are still no commonly-accepted approaches on how to include the zero-total-event studies in meta-analyses. Both views are hotly contested in the environment of clinical trials in drug safety analysis, where zero-total-event studies are common occurrences (cf. Finkelstein and Levin, 2012). Throughout the note, the word "exclusion" does not refer to physically removing the zero-total-event studies from analysis, but to using methods which yield the same results whether or not the zero–total-event studies are included in the data set. Similarly, its antonym "inclusion" is used in the same fashion.

There are two fundamental questions underlying this debate. First, does a zero-total event study in the meta-analysis setting, such as the one in Nissen and Wolski (2007), indeed contain no information about the parameter common odds ratio? If this is indeed the case, the debate would be over. If, on the contrary, a zero-total-event study is found to contain information about the common odds ratio, a natural follow-up question would be how zero-total-event studies can be effectively included in a meta-analysis. In this note, under the common assumption that the underlying true event rates are non-zero, we answer the first question. An attempt to answer the second question using an approach of combining confidence distributions and related discussions can be found in Liu et al. (2014); Yang et al. (2016). Specifically, we show in this note that the popular conditional likelihood argument is not appropriate for answering the first question, even though it provides a valid (but conservative) inference for the common odds ratio. In addition, by carefully studying the regular likelihood likelihood approach, we demonstrate that a zero-total-event study in fact contains information even for the parameter of common odds ratio under the common assumption that the underlying true event rates are non-zero. Thus it should be included in meta-analysis of two-by-two tables.

The rest of the article is arranged as follows. In Section 2, we consider the standard setup of meta-analysis of two-by-two tables where the underlying events rates are assumed non-zero, and describe approaches and issues related to rare event studies. In Section 3, we compare the regular and conditional likelihood approaches, first using a simple example of a special case, and then in general forms to highlight the difference of the two approaches. The comparisons show that zero-total event studies contain information about the common odds ratio. In Section 4, we present simulation studies to provide further support for our conclusion. Section 5 contains more discussions and remarks.

2. TWO-BY-TWO TABLES, COMMON ODDS RATIO AND META-ANALYSIS OF RARE EVENT STUDIES

Consider K independent studies with two arms, treatment versus control: $X_i \sim Binomial(n_i, \pi_{1i})$ and $Y_i \sim Binomial(m_i, \pi_{0i})$, for $i = 1, \ldots, K$, where both the event rates are non-zero $\pi_{1i} > 0$ and $\pi_{0i} > 0$. Assume that we are interested in the odds ratio $\theta_i = \{\pi_{1i}/(1-\pi_{1i})\}/\{\pi_{0i}/(1-\pi_{0i})\}$. Under the common odds ratio assumption, the θ_i's are assumed to be a constant θ across all K studies, although the rates (π_{1i}, π_{0i}) may or may not be the same from one study to another. The sum of total numbers of events in the i-th study is denoted by $T_i = X_i + Y_i$. This setup is the classical common odds ratio fixed-effect setup, in which we allow the (fixed) event rates to possibly vary from one study to another (cf., e.g., Breslow, 1981; Cox, 1989; Nissen and Wolski, 2007; Finkelstein and Levin, 2012; Tian et al., 2009, among others). Note that, by allowing the (fixed) event rates to be different in different studies, this fixed-effects model assumption is weaker than a random-effects model assumption requiring the unknown event rates to be realizations from a single distribution; see also, e.g., Clagget et al. (2014) for such a discussion.

The sample collected under such a setting is often expressed in a sequence of two-by-two tables:

(1)

x_i	$n_i - x_i$	n_i
y_i	$m_i - y_i$	m_i
t_i	$(n_i + m_i) - t_i$	$n_i + m_i$

for $i = 1, 2, \ldots, K$,

where x_i and y_i are the observed numbers of events in the treatment and control arms of the i-th study, respectively, $t_i = x_i + y_i$ and $\{x_i, y_i, t_i\}$ are a sample realization of their random counterparts $\{X_i, Y_i, T_i\}$. Let $\mathbf{X} = \{X_1, X_2, \ldots, X_K\}$, $\mathbf{Y} = \{Y_1, Y_2, \ldots, Y_K\}$ and $\mathbf{T} = \{T_1, T_2, \ldots, T_K\}$, and their realizations $\mathbf{x} = \{x_1, x_2, \ldots, x_K\}$, $\mathbf{y} = \{y_1, y_2, \ldots, y_K\}$ and $\mathbf{t} = \{t_1, t_2, \ldots, t_K\}$, respectively. In rare event studies, the observed x_i, y_i and t_i are usually very small. Those studies with $t_i = 0$ (i.e., both $x_i = 0$ and $y_i = 0$) are referred to as *zero-total-event studies*. Note that, when, $x_i = 0$ or $y_i = 0$ or both $x_i = y_i = 0$, the sample version of the odds ratio $\hat{\theta}_i = \{x_i/(n_i - x_i)\}/\{y_i/(m_i - y_i)\}$ involves (1/0) or (0/0), and thus is undefined.

There are two independent binomial random variables X_i and Y_i in the i-th two-by-two table. The likelihood function of the i-th table is simply

$$(2)\quad L_{(x_i,y_i)}(\pi_{1i}, \pi_{0i}) = \binom{n_i}{x_i}\binom{m_i}{y_i}\pi_{1i}^{x_i}(1-\pi_{1i})^{n_i-x_i}\pi_{0i}^{y_i}(1-\pi_{0i})^{m_i-y_i}.$$

The joint likelihood function across all K tables is

$$(3)\qquad L_{(\mathbf{x},\mathbf{y})}(\pi_1, \pi_0) = \prod_{i=1}^{K} L_{(x_i,y_i)}(\pi_{1i}, \pi_{0i}),$$

where $\pi_1 = (\pi_{11}, \ldots, \pi_{1K})^T$ and $\pi_0 = (\pi_{01}, \ldots, \pi_{0K})^T$. Under the common odds ratio assumption, (π_{1i}, π_{0i}) satisfy a constraint $\{\pi_{1i}/(1-\pi_{1i})\}/\{\pi_{0i}/(1-\pi_{0i})\} = \theta$. In this

case, we can re-express the joint likelihood function and the likelihood function in (2) and (3) as $L_{(\mathbf{x},\mathbf{y})}(\theta,\pi_0)$ and $L_{(x_i,y_i)}(\theta,\pi_{0i})$, respectively, with π_{1i} in the right-hand side of (2) being replaced by $\pi_{1i}=(\theta\pi_{0i})/(1-\pi_{0i}+\theta\pi_{0i})$. The parameters $\pi_0=(\pi_{01},\ldots,\pi_{0K})^T$ are considered as nuisance parameters when we carry out meta-analysis inference on θ.

Sometimes, for example in Fisher exact tests, an inference is made conditional on the given marginal total T_i. Conditional on $T_i=t_i$, there is only one random variable (cell) in the two-by-two table, and the conditional distribution of X_i given $T_i=t_i$ follows a noncentral hypergeometric distribution:

$$P_\theta(X_i=x_i|T_i=t_i) = \binom{n_i}{x_i}\binom{m_i}{t_i-x_i}\theta^{x_i}\Big/\sum_{v=a_i}^{b_i}\binom{n_i}{v}\binom{m_i}{t_i-v}\theta^{v}, \tag{4}$$

for $a_i\le x_i\le b_i$, where $a_i=\max(0,t_i-m_i)$ and $b_i=\min(n_i,t_i)$. This is also the conditional likelihood function of the i-th two-by-two table

$$\tilde{L}_{x_i|t_i}(\theta)=P_\theta(X_i=x_i|T_i=t_i) \tag{5}$$

and the joint conditional likelihood function for all K tables is

$$\tilde{L}_{\mathbf{x}|\mathbf{t}}(\theta)=\prod_{i=1}^{K}\tilde{L}_{x_i|t_i}(\theta). \tag{6}$$

The conditional likelihood function involves only the parameter of interest θ and not nuisance parameters. This makes the inference based on the conditional likelihood a much easier task.

In the context of meta-analysis of two-by-two tables, one makes inference on θ using information across all K tables. Generally speaking, a meta-analysis combines the results from multiple studies to reach an overall conclusion, and in practice typically increases statistical accuracy or power of inference. It has become a well-established and increasingly important tool in medical research and other fields. Many meta-analysis approaches have been developed and can be applied to combine information from trials summarized by two-by-two tables. They include the so-called model-based methods (including both fixed and random effects models and also Bayesian hierarchical models), the combining p-values methods, and methods developed specifically for combining two-by-two tables such as the Mantel–Haenszel and the Peto methods, among others. In this article, we are specifically interested in the case of rare events where the data in the two-by-two tables are sparse. When the data are sparse, a single study is inadequate for drawing a reliable conclusion. But conclusions can often be strengthened by using meta-analysis to synthesize findings from a number of similar studies. A challenging case, which happens often in clinical studies concerning drug safety, is that a non-negligible, sometimes even substantial, portion of the studies are zero-total-event studies. In this case, the sample odds ratio $\hat{\theta}_i=\{x_i/(n_i-x_i)\}/\{y_i/(m_i-y_i)\}$ is undefined and the true parameter values of (π_{1i},π_{0i}) are very close to the boundary value 0 (although it is still a common assumption that (π_{1i},π_{0i}) are not equal to zero). For example, the event rates for myocardial infarction in the Avandia studies are around 0.5%, which is close but not equal to zero, and as a result a number of studies are zero-total-event studies; cf., Finkelstein and Levin (2012). The debate has been on whether or not we should include these observed zero-total-event studies in meta-analysis.

In the case of rare events with the true values of (π_{1i},π_{0i}) very close to the boundary value 0 but not equal to 0, methods based on large sample asymptotics generally do not apply. The reason is that, with both (π_{0i},π_{1i}) not equal to 0, the probability of observing a zero-total-event study is 0 when $n_i\to\infty$ and $m_i\to\infty$. When a zero-total-event study is observed, it is an indication that the sample sizes are not large enough for this particular underlying set. Many conventional approaches for two-by-two tables that have been used in practice, including Mantel–Haenszel and Peto methods, either simply excluding zero-total-event studies from analysis (e.g., Nissen and Wolski, 2007) or adding an arbitrary number (often 0.5 or 0.1) to the zero cells (e.g., Diamond et al., 2007), are justified based on large sample asymptotics. Their practical results in meta-analysis of two-by-two tables of rare events are mixed. Recently, Cai et al. (2010) studied the performance of model based meta-analysis approaches on two-by-two tables. But again the methods rely on asymptotic justifications. Tian et al. (2009) proposed an exact approach for meta-analysis of risk difference by combining a sequence of confidence intervals. Although the method by Tian et al. (2009) does not rely on large sample asymptotics, it cannot handle the parameter of odds ratio, because the sample version $\hat{\theta}_i$ is undefined unless the zero-total-event studies are excluded *a priori*.

For the parameter odds ratio, conditional likelihood inference approach has also been proposed for meta-analysis of two-by-two tables, which is considered exact inference; see, e.g., Cox (1989). In particular, one can make an inference relying on the joint conditional likelihood function $\tilde{L}_{\mathbf{x}|\mathbf{t}}(\theta)$ defined in (6). Mehta et al. (1985) provided a computing algorithm to perform exact inference based on the joint conditional likelihood function (6). Davison (1988) provided a small sample approximation of the conditional likelihood based on saddlepoint approximations, which leads to approximate inference with much less computing effort. Under the conditional inference framework, the conditional likelihood function of a zero-total-event study is $\tilde{L}_{x_i=0|t_i=0}(\theta)\equiv 1$, and thus the study does not contribute to the inference of θ. This is perhaps the most prominent argument for excluding zero-total-event studies in meta-analysis, because ultimately the zero-total-studies do not have impact on the conditional likelihood function. An attractive feature

of the conditional likelihood inference is that it does not depend on nuisance parameters, which makes the analysis an easier task.

However, in most clinical trials of safety studies such as the case in Nissen and Wolski (2007), only the numbers of treatments and controls (n_i, m_i) can be viewed as preset. The sample realizations t_i for the sums of the numbers of events in both treatment and control arms are not fixed in advance. In the next two sections (Sections 3 and 4), we use simple numerical examples as well as comparisons of regular and conditional likelihood functions in the general setting, to show that (i) the argument based on conditional likelihood functions, although valid in terms of ensuring the size of the test (Lydersen et al., 2009), is conservative and thus suffers loss of information and (ii) zero-total-event studies actually contain information about the common odds ratio θ.

3. LIKELIHOOD VERSUS CONDITIONAL LIKELIHOOD

3.1 A simple illustrative example

We begin with a simple example which shows that the inferences based on the regular likelihood and the inference based on the conditional likelihood are very different in meta-analyses of two-by-two tables. Note that, in the general setup of two-by-two tables (1), we assume only that the odds ratios are the same; the underlying binomial rates (π_{0i}, π_{1i}) are not zero and they may or may not be the same from one study to another. For simplicity, we further assume in this simple example that the underlying event rates (π_{0i}, π_{1i}) are also the same across the K studies, which is a special case of the setup (1). This simple example suffices to make our point without clouding the comparison with other complications. The mathematical insight behind this simple example can be generalized to the general situations.

Example 1 Suppose that we have a total of 40 individuals, 20 of them receive a treatment and the other 20 are the controls. The event rate for the treatment is π_1 and the rate for the control is π_0. The 40 individuals are randomly assigned to 10 studies of sizes $n_i = 2$ versus $m_i = 2$. We further assume that we observe a sample realization of three two-by-two tables of type-A and seven tables of type-B as shown below:

(7)

1	1	2		0	2	2
0	2	2		0	2	2
1	3	4		0	4	4
type-A				type-B		

A type-B table is a zero-total-event study. From (2), the corresponding likelihood function of the type-A tables is $2\pi_1(1-\pi_1)(1-\pi_0)^2$ and the likelihood function of the type-B tables is $(1-\pi_1)^2(1-\pi_0)^2$. Thus, the joint likelihood function is

(8)
$$L^{[\text{split}]}_{(\mathbf{x},\mathbf{y})}(\theta,\pi_0) = 2^3\pi_1^3(1-\pi_1)^3(1-\pi_0)^6(1-\pi_1)^{14}(1-\pi_0)^{14} = 8\pi_1^3(1-\pi_1)^{17}(1-\pi_0)^{20},$$

where $\pi_1 = (\theta\pi_0)/(1-\pi_0+\theta\pi_0)$. Note that the likelihood function of the type-B zero-total-event study is $(1-\pi_0)^2(1-\pi_1)^2 = (1-\pi_0)^4/(1-\pi_0+\theta\pi_0)^2$, which is not a constant and depends on both π_0 and θ. Such studies contribute to the inference on both θ and π_0 when we use the regular likelihood function.

Suppose now that we pool the 10 two-by-two tables into a single table for all 40 individuals:

(9)

3	17	20
0	20	20
3	37	40

.

The likelihood function corresponding to the pooled table is

$$L^{[\text{whole}]}_{(\mathbf{x},\mathbf{y})}(\theta,\pi_0) = 1140\pi_1^3(1-\pi_1)^{17}(1-\pi_0)^{20}, \tag{10}$$

which is the same as $L^{[\text{split}]}_{(\mathbf{x},\mathbf{y})}(\theta,\pi_0)$ in (8), up to a constant. Clearly, the inference based on these two versions of regular likelihood functions does not change, regardless whether the data are pooled in one table or separated into several independent tables. This is not surprising, since in this example the 20 individual samples from either the control or the treatment group are i.i.d. samples, and our inference should not depend on how we split the overall table into several small tables. In this case, regardless whether we present the data in a single table or randomly separate them into 10 tables, the information contained in the data does not change. This statement is backed up by the joint likelihood functions of the full parameters (θ,π_0) that are the same $L^{[\text{split}]}_{(\mathbf{x},\mathbf{y})}(\theta,\pi_0) \equiv L^{[\text{whole}]}_{(\mathbf{x},\mathbf{y})}(\theta,\pi_0)$, up to a constant.

Now, let us study the behavior of the conditional likelihood inference, which turns out to be quite different. Based on equation (4), the conditional likelihood function of the type-A tables is $\log\theta/(1+\log\theta)$ and the conditional likelihood function of the type-B tables is the constant 1. So the joint conditional likelihood function of the 10-table sample realization in (7) is

$$\tilde{L}^{[\text{split}]}_{\mathbf{x}|\mathbf{t}}(\theta) = \theta^3/(1+\theta)^3. \tag{11}$$

But the joint conditional likelihood function of the pooled table (9) is

$$\tilde{L}^{[\text{whole}]}_{\mathbf{x}|\mathbf{t}}(\theta) = \theta^3 \Big/ \left(1+\frac{9}{2}\theta+\frac{9}{2}\theta^2+\theta^3\right). \tag{12}$$

Clearly, (11) and (12) are different. In addition, if we ignore the seven zero-total-event studies of the type-B tables in (7) and pool the three not-zero-total-event studies of the type-A tables, we obtain a two-by-two table:

(13)
$$\begin{array}{cc|c} 3 & 3 & 6 \\ 0 & 6 & 6 \\ \hline 3 & 9 & 12 \end{array}.$$

The conditional likelihood function of table (13) is

$$\tilde{L}^{[\text{whole.no.zero}]}_{\mathbf{x}|\mathbf{t}}(\theta) = e^{3\theta} \Big/ \left(1 + \frac{10}{3}\theta + \frac{10}{3}\theta^2 + \theta^3\right).$$

These three conditional likelihood functions $\tilde{L}^{[\text{split}]}_{\mathbf{x}|\mathbf{t}}(\theta)$, $\tilde{L}^{[\text{whole}]}_{\mathbf{x}|\mathbf{t}}(\theta)$ and $\tilde{L}^{[\text{whole.no.zero}]}_{\mathbf{x}|\mathbf{t}}(\theta)$ are all different. This is not surprising, since these conditional likelihood functions depend on their respective table margins.

In this special example of 20 i.i.d samples from $Bernoulli(\pi_0)$ and 20 i.i.d samples from $Bernoulli(\pi_1)$, it is desirable to reach the same inference conclusions regardless of whether we split them into separate tables or not. The regular likelihood approaches support this statement exactly. But, in comparison, the conditional likelihood approach would reach different conclusions depending on different splitting and table margins.

In a more general setup, the event rates (π_{i0}, π_{i1}) may be different from one study to another (even though the odds ratio are identical). In this case, $L^{[\text{split}]}_{(\mathbf{x},\mathbf{y})}(\theta, \pi_0)$ and $L^{[\text{whole}]}_{(\mathbf{x},\mathbf{y})}(\theta, \pi_0)$ may be different. The appropriate inference should be based on $L^{[\text{split}]}_{(\mathbf{x},\mathbf{y})}(\theta, \pi_0)$ and it is not appropriate to pool all 40 individual results into one single table and use $L^{[\text{whole}]}_{(\mathbf{x},\mathbf{y})}(\theta, \pi_0)$, since $\pi_{10}, \ldots, \pi_{K0}$ do not share the same value and pooling introduces bias. Note that the difference between $L^{[\text{split}]}_{(\mathbf{x},\mathbf{y})}(\theta, \pi_0)$ and $L^{[\text{whole}]}_{(\mathbf{x},\mathbf{y})}(\theta, \pi_0)$ in this case is caused by the change of event rates (π_{i0}, π_{i1}) from one table to another. This is not the same as the cause of the three different results in the conditional likelihood inference in the illustrative example discussed in this subsection – the different realizations of the table margins t_i. In the more general setup with the event rates (π_{i0}, π_{i1}) being different from one study to another, the different realizations of the table margins t_i still affect conditional likelihood inferences, in addition to the change of event rates (π_{i0}, π_{i1}) of different tables. In any case, it is not appropriate to use either $\tilde{L}^{[\text{whole}]}_{\mathbf{x}|\mathbf{t}}(\theta)$ or $\tilde{L}^{[\text{whole.no.zero}]}_{\mathbf{x}|\mathbf{t}}(\theta)$. Although an inference based on $\tilde{L}^{[\text{split}]}_{\mathbf{x}|\mathbf{t}}(\theta)$ still provides a valid inference (in terms of ensuring the size of the test) as shown in Lydersen et al. (2009), it is conservative and not as efficient as $L^{[\text{split}]}_{(\mathbf{x},\mathbf{y})}(\theta, \pi_0)$. This point will be further elaborated in the remainder of this section and also the remainder of this paper.

3.2 A difference between the regular likelihood function and the conditional likelihood function

We now proceed in the general setup of the given design in Section 2 to examine further the difference between the joint regular likelihood function $L_{(\mathbf{x},\mathbf{y})}(\theta, \pi_0)$ in (3) and the conditional likelihood function $\tilde{L}_{(\mathbf{x}|\mathbf{t})}(\theta)$ in (6). We use here the *strong likelihood principle*, namely, all information from the data relevant to the inference is contained in the regular likelihood function (cf., e.g., Berger and Wolpert 1988). Keep in mind that both the underlying true event rates are non-zero $\pi_{1i} > 0$ and $\pi_{0i} > 0$.

Since $\prod_{i=1}^{K} P(X_i = x_i, T_i = t_i) = \prod_{i=1}^{K} P(X_i = x_i | T_i = t_i) \prod_{i=1}^{K} P(T_i = t_i)$, it follows immediately that

$$L_{(\mathbf{x},\mathbf{y})}(\theta, \pi_0) = \tilde{L}_{(\mathbf{x}|\mathbf{t})}(\theta) D_{\mathbf{t}}(\theta, \pi_0), \tag{14}$$

where the ratio difference of the regular and conditional likelihood functions $L_{(\mathbf{x},\mathbf{y})}(\theta, \pi_0) / \tilde{L}_{(\mathbf{x}|\mathbf{t})}(\theta)$ is

$$\begin{aligned} D_{\mathbf{t}}(\theta, \pi_0) &= \prod_{i=1}^{K} P(T_i = t_i) \\ &= \prod_{\{i:\, t_i \neq 0\}} P(T_i = t_i) \prod_{\{i:\, t_i = 0\}} P(T_i = t_i) \\ &= \prod_{\{i:\, t_i \neq 0\}} \sum_{a=0}^{t_i} \binom{n_i}{a} \pi_{1i}^{a} (1 - \pi_{1i})^{n_i - a} \\ &\quad \times \binom{m_i}{t_i - a} \pi_{0i}^{t_i - a} (1 - \pi_{1i})^{m_i - t_i + a} \\ &\quad \times \prod_{\{i:\, t_i = 0\}} (1 - \pi_{1i})^{n_i} (1 - \pi_{0i})_i^{m}. \end{aligned}$$

In the special example studied in Section 3.1, the term $D_{\mathbf{t}}(\theta, \pi_0) = L^{[\text{split}]}_{(\mathbf{x},\mathbf{y})}(\theta, \pi_0) / \tilde{L}^{[\text{split}]}_{(\mathbf{x}|\mathbf{t})}(\theta) = \big\{(1-\pi_1)^2 \pi_0 (1-\pi_0)^1 + \pi_1 (1-\pi_1)^2 (1-\pi_0)^2\big\}^3 \big\{(1-\pi_1)^2 (1-\pi_0)^2\big\}^7 = (1-\pi_1)^{17} (1-\pi_0)^{17} \big\{(1-\pi_1)\pi_0 + \pi_1 (1-\pi_0)\big\}^3$. This term, as a function of θ and π_0, is plotted in Figure 1 (a). The plot clearly shows that the ratio difference $D_{\mathbf{t}}(\theta, \pi_0)$ depends on both π_0 and θ. So, the conditional inference conditional on the marginal total T_i, although simple for making inference of the parameter of interest θ, typically is different than that based on the regular likelihood function. Following the strong likelihood principle that all information is contained in the regular likelihood function, the difference would suggest that the conditional likelihood approach can incur omission or distortion of information (even though such an omission or distortion could be minor in some situations).

Furthermore, the contribution of the total-zero-event studies to the joint regular likelihood function is

$$C_0(\theta, \pi_0) = \prod_{\{i:\, t_i = 0\}} (1 - \pi_{1i})^{n_i} (1 - \pi_{0i})^{m_i}.$$

In the special example studied in Section 3.1, this term is $C_0(\theta, \pi_0) = (1-\pi_1)^{14} (1-\pi_0)^{14}$ and we plot it as a function of π_0 and θ in Figure 1(b). Again, $C_0(\theta, \pi_0)$ clearly depends on both π_0 and θ too. However, this part of information from

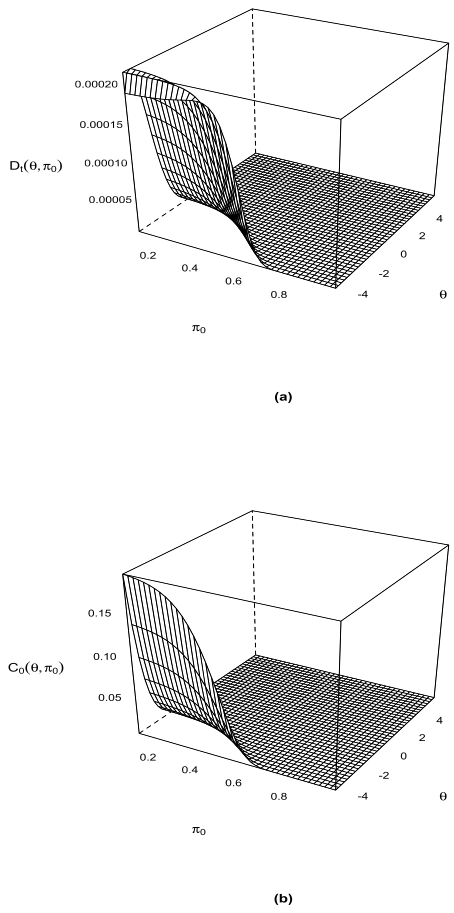

Figure 1. Plots of (a) $D_{\mathbf{t}}(\theta,\pi_0)$ and (b) $C_0(\theta,\pi_0)$, as functions of π_{0i} and θ.

$C_0(\theta,\pi_0)$ is not utilized in the conditional likelihood inference for θ. In other words, the conditional inference may distort or omit information about θ. Consequently, conclusions reached under conditional likelihood inference where the zero-total-event studies are ignored may suffer loss of efficiency.

Finally, we examine Fisher information under the general setting of two-by-two tables (1). The Fisher information matrix is

$$\mathbf{I}_n = -\mathrm{E}\left\{\frac{\partial^2}{\partial(\theta,\pi_0)^T\partial(\theta,\pi_0)}\log L_{(\mathbf{x},\mathbf{y})}(\theta,\pi_0)\right\}.$$

We denote by the information matrix by the conditional inference method given $\mathbf{t}=(t_1,\ldots,t_K)$,

$$\tilde{\mathbf{I}}_n(\mathbf{t}) = -\mathrm{E}\left\{\frac{\partial^2}{\partial\theta^2}\log \tilde{L}_{(\mathbf{x}|\mathbf{t})}(\theta)\middle|\mathbf{t}\right\}.$$

Based on (14) and a direct calculation, we can relate $\mathbf{I}_n(\mathbf{t})$ to $\mathbf{I}_n$:

$$\mathbf{I}_n = \begin{pmatrix}\mathrm{E}\tilde{\mathbf{I}}_n(\mathbf{t}) & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix} + \sum_{i=1}^{K}\mathrm{E}\left\{-\frac{\partial^2}{\partial(\theta,\pi_0)^T\partial(\theta,\pi_0)}\log P(T_i=t_i)\right\} \geq \begin{pmatrix}\mathrm{E}\tilde{\mathbf{I}}_n(\mathbf{t}) & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}$$

The $(1,1)$ element of $\mathbf{I}_n$ is

(15)
$$\mathbf{I}_n^{(1,1)} = \mathrm{E}\tilde{\mathbf{I}}_n(\mathbf{t}) + \sum_{i=1}^{K}\mathrm{E}\left\{-\frac{\partial^2}{\partial\theta^2}\log P(T_i=t_i)\right\} \geq \mathrm{E}\tilde{\mathbf{I}}_n(\mathbf{t}),$$

and the strict inequality holds typically when $\pi_{1i}>0$ and $\pi_{0i}>0$ for all i. Apparently, the information matrices $\tilde{\mathbf{I}}_n(\mathbf{t})$ and $\mathbf{I}_n$ are different and, when average over all possible $\mathbf{t}$ margins, $\mathbf{I}_n^{(1,1)}-\tilde{\mathbf{I}}_n(\mathbf{t})$ is greater than zero.

In particular, the information difference $\mathbf{I}_n^{(1,1)}-\tilde{\mathbf{I}}_n(\mathbf{t})$ contributed by a single study is

(16)
$$\mathrm{E}_T\left\{-\frac{\partial^2}{\partial\theta^2}\log P(T=t)\right\} = \sum_{t=0}^{n+m}\left\{-\frac{\partial^2}{\partial\theta^2}\log D_t(\theta,\pi_0)\right\}\cdot D_t(\theta,\pi_0).$$

To numerically evaluate this difference, we let $\psi=1-\pi_0+\theta\pi_0$ and write

$$D_t(\theta,\pi_0) = \frac{\pi_0^t(1-\pi_0)^{n+m-t}}{\psi^n}\sum_{a=0}^{t}\binom{n}{a}\binom{m}{t-a}\theta^a.$$

Therefore

$$-\frac{\partial^2}{\partial\theta^2}\log D_t(\theta,\pi_0) = -\frac{\frac{\partial^2}{\partial\theta^2}D_t(\theta,\pi_0)}{D_t(\theta,\pi_0)} + \left\{\frac{\frac{\partial}{\partial\theta}D_t(\theta,\pi_0)}{D_t(\theta,\pi_0)}\right\}^2,$$

where

$$\frac{\partial}{\partial\theta}D_t(\theta,\pi_0) = \frac{\pi_0^t(1-\pi_0)^{n+m-t}}{\psi^n}\sum_{a=0}^{t}\binom{n}{a}\binom{m}{t-a}\left(a-\frac{n\pi_0}{\psi}\right)\theta^{a-1}$$

and

$$\frac{\partial^2}{\partial\theta^2}D_t(\theta,\pi_0) = \frac{\pi_0^t(1-\pi_0)^{n+m-t}}{\psi^{n+2}}\times \sum_{a=0}^{t}\binom{n}{a}\binom{m}{t-a}\left\{(a\psi-n\pi_0\theta)^2-a\psi^2+n\pi_0^2\theta^2\right\}\theta^{a-2}.$$

As an illustration, we set $n=m=2$ in the setting of Example 1 and examine the size of (16) when varying values of θ and π_0. The numerical result in Part I of Table 1 suggests that for any given value of the odds ratio θ, the information loss by the conditional inference is always positive, but it tends to be less when the baseline probability π_0 is getting smaller. A similar pattern is observed in Part II of Table 1 where $n=m=100$.

Table 1. Part I. Numerical evaluation of the information difference (16) when $n = m = 2$

	Odds ratio θ								
π_0	1/5	1/4	1/3	1/2	1	2	3	4	5
0.1	0.97	0.62	0.36	0.19	0.09	0.05	0.04	0.03	0.03
0.2	4.18	2.48	1.26	0.50	0.16	0.09	0.07	0.06	0.04
0.3	11.04	6.30	2.96	0.98	0.21	0.12	0.09	0.07	0.06
0.4	23.96	13.22	5.85	1.67	0.24	0.14	0.11	0.09	0.07
0.5	47.31	25.25	10.53	2.63	0.25	0.16	0.13	0.10	0.08

Part II. Numerical evaluation of the information difference (16) when $n = m = 100$

	Odds ratio θ								
π_0	1/5	1/4	1/3	1/2	1	2	3	4	5
0.01	17.11	9.95	4.81	1.68	0.49	0.57	0.66	0.71	0.74
0.02	67.83	38.76	17.98	5.43	0.98	1.58	2.09	2.34	2.47
0.05	438.62	247.23	111.03	29.74	2.37	7.06	10.12	11.31	11.62
0.10	1900.53	1061.38	467.76	117.91	4.50	22.98	32.27	34.22	33.31
0.20	9095.37	5005.49	2149.83	508.91	8.00	72.67	92.35	88.84	79.61

3.3 Frequentist interpretations of regular and conditional likelihood approaches

In this subsection, we use frequentist interpretations to explain the regular and conditional likelihood approaches, respectively. In these interpretations, we highlight the conditions under which the regular and conditional likelihood inference are developed. The comparison provides yet another angle and an additional reason that the conditional likelihood approach is different than the regular likelihood approach. It highlights that the conditional likelihood approach is a valid but conservative method for the meta-analysis considered in this article.

Consider two types of probability formulations, which we refer to as *the unconditional probability formulation* if our probability statements are based on random sample $(\mathbf{X}, \mathbf{Y})$; and as *the conditional probability formulation* if our probability statements are based on the conditional distribution of $(\mathbf{X}, \mathbf{Y})$ conditional on fixed (realized) margins $\mathbf{T} = \mathbf{t}$. We interpret results from the regular and conditional likelihood approaches based on these two different probability formulations.

- **Regular likelihood approach under the unconditional probability formulation** When we apply the regular likelihood approach to two-by-two tables to draw inference, say Statement S, our probability formulation is based on random sample $(\mathbf{X}, \mathbf{Y})$. So it is understood that this inference is done based on K independent Binomial distributions where only side margins (n_i, m_i) are preset. Moreover, using the frequentist interpretation and the standard 95% confidence level as an example, this inference procedure is interpreted as *if one repeats the experiment under the same Binomial setting (with only fixed (n_i, m_i)'s)*, Statement S is correct 95% of the time.
- **Conditional likelihood approach under the conditional probability formulation** When we apply the conditional likelihood approach to two-by-two tables to draw inference, say Statement S', our probability formulation is based on the conditional distribution of $(\mathbf{X}, \mathbf{Y})$ conditional on fixed (realized) margins $\mathbf{T} = \mathbf{t}$. A strict interpretation of this conditional probability formulation should follow the fact that our statement is obtained conditioning on the observed (a *specific* set of) margins t_i, in addition to the preset margins (n_i, m_i). Based on this fact and using the standard 95% confidence level as an example, the frequentist interpretation of the result from the conditional likelihood inference procedure is then: *if one repeats the experiment under the exact same setting (including both the fixed specific table margins t_i and the margins (n_i, m_i))*, Statement S' is correct 95% of the time.
- **Conditional likelihood approach under the unconditional probability formulation** For a hypothesis testing problem, say with a null hypothesis H_0, Lydersen et al. (2009) showed that any conditional test (developed assuming that both (n_i, m_i) and t_i are fixed) preserves test size for the setting of only fixing side margins (n_i, m_i) but not the side margins of T_i,. The justification is based on the following inequality:

$$
\begin{aligned}
&P_{(X_i,Y_i)}(\text{Reject } H_0|H_0) \\
&= \sum_{t_i} P_{(X_i,Y_i)|T_i=t_i}(\text{Reject } H_0|t_i, H_0) P_{T_i}(t_i|H_0) \\
&\le \sum_{t_i} \alpha P_{T_i}(t_i|H_0) = \alpha \sum_{t_i} P_{T_i}(t_i|H_0) = \alpha.
\end{aligned} \tag{17}
$$

Here, $P_{(X_i,Y_i)}(\cdot|H_0)$ is the joint probability statement on (X_i, Y_i), $P_{(X_i,Y_i)|T_i=t_i}(\cdot|t_i, H_0)$ is the conditional probability statement on (X_i, Y_i) given $T_i = t_i$ and $P_{T_i}(\cdot|H_0)$ is the marginal probability statement on T_i, assuming H_0 is true and for each i. A similar statement can be developed for confidence intervals, due to the well-known duality between tests and confidence sets in

which one can be derived from the other and vice-versa. Thus, a conditional likelihood approach can also be justified under the general unconditional probability formulation: *if one repeats the experiment under the same Binomial setting (with only fixed* (n_i, m_i)*'s)*, Statement S' is correct *at least* 95% of the time. Note that the validity justification of conditional inference under the unconditional probability formulation is based on inequality (17). The use of such an inequality spells out exactly the conservative nature of a conditional inference. This message matches with what we have reached by directly comparing the regular and conditional likelihood functions in the earlier sections. The same message was also presented in the simulation studies by Lydersen et al. (2009).

For a clinical trial, one typically presets the total sample sizes n_i and m_i in treatment and control, but definitely not the marginal total $t_i = x_i + y_i$ in any table. So for an approach to be valid under the typical meta-analysis setting of clinical trials, it needs to be justified under the interpretation of the unconditional probability formulation stated above. Clearly, the underlying requirement in the regular likelihood approach matches the typical clinical trial setup and thus justifies the approach. On the other hand, the underlying requirement in the conditional likelihood approach does not match the typical setting of clinical trials, even though it can be justified as a valid approach in the unconditional setting using inequality (17). Such reliance on inequalities subsequently causes the conditional likelihood approach to lose power. Again, we cannot rely on a conditional likelihood statement to determine whether a zero-total-event study contains any information about the common odds ratio θ or any other parameters.

4. SIMULATION STUDY

We conduct simulation studies to numerically examine the impact of zero-total-event studies. In accordance with the theoretical examinations in earlier sections, we consider the setting of meta-analysis of independent two-by-two tables, assuming that each table (study) shares a common odds ratio and its event rates are non-zero. We consider the two-sided hypothesis testing problem $H_0 : \theta = 1$ versus $H_1 : \theta \neq 1$ for the common odds ratio. We perform analysis using the conditional likelihood method and also several variants of likelihood ratio tests based on the regular likelihood function.

Specifically, we compare analysis results obtained from the following approaches:

1. *The conditional likelihood approach.* This approach conditions on both the row totals (n_i, m_i) and the column totals $(t_i, n_i + m\_t_i)$ in each table (1). Given both the row and column totals, the table layout solely relies on X_i which follows the hypergeometric distribution (4). Therefore, exact inference can be drawn from the exact distribution of $\sum_{i=1}^{K} X_i$, conditional on all the marginal totals of the 2×2 tables (Mehta, Patel, and Gray, 1985). We implement the analysis using SAS PROC FREQ with specification of the COMOR option in the EXACT statement. The computing of this exact method is based on Vollset, Hirji, and Elashoff (1991). In this approach, zero-total-event studies do not have impact on the inference and are in effect excluded in the meta-analysis.
2. *The likelihood ratio test (LRT) approaches.* The likelihood ratio test statistic is

$$Z_{LR} = 2 \left\{ \max_{(\pi_{1i}, \pi_{0i},\ i=1,\ldots,K)} \sum_{i=1}^{K} \log L(\pi_{1i}, \pi_{0i}) - \max_{H_0} \sum_{i=1}^{K} \log L(\pi_{1i}, \pi_{0i}) \right\}, \tag{18}$$

where $L(\pi_{1i}, \pi_{0i})$ has the expression as shown in (2). The calculation of the test statistic z_{LR} involves K unknown nuisance parameters π_{0i}'s. In addition, the null distribution of Z_{LR} does not have a closed form. Depending on how we estimate the nuisance parameters and how we simulate data under the null so as to render the null distribution of Z_{LR}, we consider the following three variants of the LRT approach:

   – The LRT-true-MC method uses the true values of the nuisance parameters π_{0i} for calculating the test statistic z_{LR} and also for simulating data under the null to render the null distribution of Z_{LR} in (18). Specifically, under the null hypothesis $H_0 : \theta = 1$ and with the true parameters π_{0i}, we repeatedly simulate $R = 10{,}000$ sets of samples and compute the corresponding Z_{LR} values. The empirical distribution of the $R = 10{,}000$ Z_{LR} values are used as an approximate to the null distribution of Z_{LR}. Since the true values of parameters are unknown, the method is applicable only in simulations. Nevertheless, it provides a reference for our study.

   – The LRT-beta-MC1 method uses a set of empirical estimates of the nuisance parameters π_{0i} for calculating the test statistic z_{LR} but use the true values of π_{0i} for simulating data to render the null distribution of Z_{LR}. To obtain the empirical estimates, we impose a working assumption that π_{0i} follow a $Beta(\beta_1, \beta_2)$ distribution. The beta distribution family is broad enough to capture or approximate a wide range of distributions with support on the interval $[0, 1]$; cf Appendix in Liu et al. (2014) for details. The unknown parameters (β_1, β_2) are estimated by maximizing the likelihood function calculated under the model with the working assumption. Since we use the true values of π_{0i} for simulating the null distribution of Z_{LR}, this method is only applicable in simulation studies.

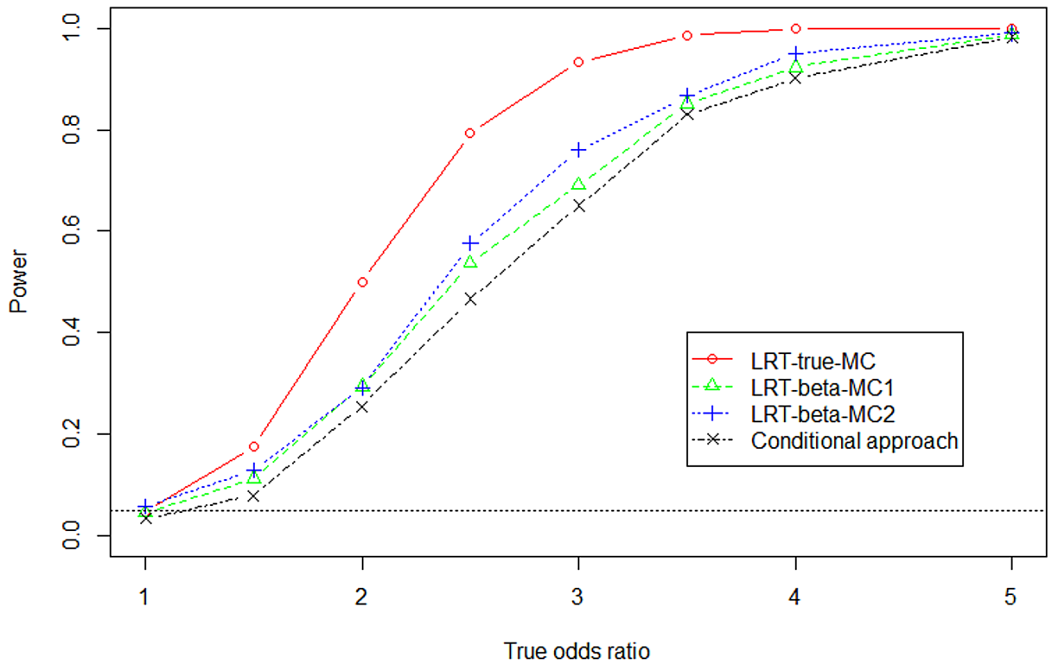

Figure 2. Comparison of power for testing the two-sided hypothesis $H_0 : \theta = 1$ *versus* $H_1 : \theta \neq 1$ *for the common odds ratio at the level* $\alpha = 0.05$. *Included are the conditional likelihood approach and the three variants of the LRT approaches described in Section 4. The results are obtained from analyzing 6 contingency tables with the event rates being* $\{0.1, 0.1, 0.05, 0.05, 0.01, 0.01\}$ *in the control arms. A dashed horizontal line at* $y = 0.05$ *is drawn for reference.*

– The LRT-beta-MC2 method uses the empirical estimates of the nuisance parameters π_{0i} mentioned in the LRT-beta-MC1 method for calculating the test statistic z_{LR} and also for simulating data to render the null distribution of Z_{LR}. Since it does not require knowledge of the true parameter values, this method can be implemented in real data applications. To save computational time, we use the Monte-Carlo method to simulate $R = 1{,}000$ sets of samples to obtain the null distribution for Z_{LR}.

For all the above three variants of the LRT approach, zero-total-event studies contribute to the likelihood and thus the test. The true values of (π_{0i}, π_{1i}) are positive by model assumption. To ensure that the empirical estimates $(\hat{\pi}_{0i}, \hat{\pi}_{1i})$ are positive in methods LRT-beta-MC1 and LRT-beta-MC2, a beta distribution is imposed on π_{0i} to serve as a catalyst to borrow information from other studies to estimate the event rates in zero-total-event studies. The numerical studies in Liu et al. (2014) showed that the empirical estimates perform much better than the simple sample mean estimates $(x_i/n_i, y_i/m_i)$ (which result in (0,0) for zero-total-event studies).

As the first part of our simulation study, we generate $K = 6$ contingency tables (1). Each of the simulated $K = 6$ tables is a summary of two independent binomial trials $X_i \sim Binomial(n_i, \pi_{1i})$ and $Y_i \sim Binomial(m_i, \pi_{0i})$, one for treatment and the other for control. In the control arms, the sample sizes are $\{25, 25, 30, 20, 30, 20\}$ and the corresponding event rates are $\{0.1, 0.1, 0.05, 0.05, 0.01, 0.01\}$. In the treatment arms, the sample sizes are $\{25, 25, 20, 30, 30, 20\}$ and the corresponding event rates are determined by $\text{logit}(\pi_{1i}) = \log(\psi) + \text{logit}(\pi_{0i})$ for a fixed odds ratio θ. This simulation setting ensures low event rates across all the studies and a non-negligible probability of observing zero-total-event studies. To understand the impact of zero-total-event studies, we analyze the simulated 2×2 tables and make inference on the common odds ratio using the four approaches described above. Specifically, over a range of θ values, we compute the power of the four approaches for testing the two-sided hypothesis $H_0 : \theta = 1$ versus $H_1 : \theta \neq 1$ at the level $\alpha = 0.05$. The results based on 1,000 simulation replicates are summarized in Figure 2.

Figure 2 shows that the conditional likelihood approach exhibits the lowest power among all the approaches under comparison. It suggests that the conditional likelihood approach bears an appreciable power loss compared to the three LRT approaches. Since the working assumption of the Beta distribution of π_{0i} is not met in our simulation setting, the latter two LRT approaches may have already suffered weakened inference power. Nevertheless, they both still have shown more power than the conditional likelihood approach. In comparison with the LRT-true-MC approaches where the true values of the nuisance parameters are used, the power loss of the conditional likelihood approach is considerable.

We note that even though the LRT-true-MC approach is not applicable in real data analysis, it provides an upper bound of the power curve for any approach.

As the second part of our simulation study, we follow the numerical studies in Tian et al. (2009) and Liu et al. (2014) and mimic the setting of the Avandia data as examined by Nissen and Wolski (2007). Specifically, we generate $K = 48$ contingency tables (1) with the sample sizes being the same as those in the Avandia data (see Table 3 in Nissen and Wolski 2007). For each of the $K = 48$ tables (studies), the event rate π_{0i} in the control arm is generated from a uniform distribution $\mathrm{U}(0, 0.01)$, and the event rate in the other arm is determined by the fixed odds ratio θ. This simulation setting ensures low event rates across all the studies and a non-negligible probability of observing zero-total-event studies. (Here, the true event rate π_{0i} is "exchangeable"; cf., e.g., Zhang et al., 2014.) The analysis results are similar to those in Figure 2 (and thus omitted), with only a change that the power of the conditional likelihood approach is closer to that of the LRT-beta approaches. This change is due to the increase of the study sample sizes (n_i, m_i) and also the number of $K = 48$ studies in the cohort. The simulation result suggests that in the rare events setting, even though n_i's and K are large, conditional inference still suffers loss of information.

5. CONCLUDING REMARKS

In this note, we are able to demonstrate that a zero-total-event study contains information for the inference on the parameters of the common odds ratio in meta-analysis when the underlying true events rates in the study are not zero. This is done by examining the difference between likelihood and conditional likelihood methods, the information matrices, and through numerical examples. We have also compared and contrasted the frequentist interpretation, as well as conditions under which the regular likelihood and conditional likelihood methods apply. The conditional likelihood method is derived under the assumption that the observed margins totals t_i are fixed (pre-specified), which is different from the typical setup of clinical trial as discussed in Nissen and Wolski (2007). The conditional likelihood method is appropriate for applications where the margins totals t_i are also pre-specified. For the trials often used in clinical studies where the margins totals t_i are not pre-specified, the validity of the conditional likelihood method can only be justified as a conservative method that ensures the correct testing size but with possible loss of power. This finding is contrary to a common belief held by many statisticians that was inappropriately justified by conditional likelihood arguments. We hope our finding here will help stop perpetuating the debate on whether or not a zero-total-event study contains information for meta-analysis.

The conclusion applies whenever we observe zero-total event studies, regardless of whether or not the event rates (π_{0i}, π_{1i}) are very small. The setting of interest, however, is rare event studies, since there is a great chance to observe zero-total event studies in rare event studies but the chance to observe such studies is small when (π_{0i}, π_{1i}) are not small. Although it is not the focus of this note, the question on how to effectively use a zero-total-event study in meta-analysis for the common odds ratio is also very important. In our simulation studies, LRT-beta-MC2 is proposed as a practical method to use the information, where a working assumption on the event rates for control π_{0i} is imposed. In literature, there are also some attempts (e.g., Liu et al. (2014); Yang et al. (2016)). However, to fully address this question, more research is needed. One may also use a Bayesian approach or a random effects model (e.g., an exchangeable model) to study these questions concerning zero-total-event studies. We can demonstrate that zero-total event studies typically still contribute to the inference in the Bayesian and analysis random effects model too. However, the use of priors imposes an additional model assumption on the problem and it "may raise more questions than they settle" (cf., Finkelstein and Levin, 2012). The same statement can be applied to a random-effects model too.

Received 5 January 2017

REFERENCES

Berger, J. and Wolpert, R. (1988), *The Likelihood Principle*, Haywood, CA: The Institute of Mathematical Statistics., 2nd ed. MR0773665

Bradburn, M. J., Deeks, J. J., Berlin, J. A., and Localio, A. R. (2007), "Much ado about nothing: A comparison of the performance of meta-analytical methods with rare events," *Statist. Med.*, 26, 53–77. MR2312699

Breslow, N. (1981), "Odds ratio estimators when the data are sparse," *Biometrika*, 68, 73–84. MR0614943

Cai, T., Parast, L., and Ryan, L. (2010), "Meta-analysis for rare events," *Statist. Med.*, 29, 2078–2089. MR2756556

Claggett, B., Xie, M., and Tian, L. (2014), "Meta analysis with fixed, unknown, study-specific parameters," *Journal of the American Statistical Association*, 109, 1667–1671. MR3293618

Cox, D. (1989), *The analysis of binary data (2nd edition)*, Chapman & Hall/CRC. MR1014891

Davison, A. C. (1988), "Approximate conditional inference in generalized linear models," *Journal of the Royal Statistical Society, Series B*, 50, 445–461. MR0970979

Diamond, G. A., Bax, L., and Kaul, S. (2007), "Uncertain effects of rosiglitazone on the risk for myocardial infarction and cardiovascular death," *Ann. Intern. Med.*, 147, 578–581.

Finkelstein, M. O. and Levin, B. (2012), "Meta-Analysis of "Sparse" Data: Perspectives From the Avandia Cases," *Jurimetrics Journal*, 52, 123–153.

Hald, A. (1952), *Statistical Theory with Engineering Applications*, New York: Wiley. MR0049514

Liu, D., Liu, R., and Xie, M. (2014), "Exact meta-analysis approach for discrete data and its application to 2×2 tables with rare events," *J. Am. Statist. Assoc.*, 109, 1450–1465. MR3293603

Lydersen, S., Fagerland, M. W., and Laake, P. (2009), "Recommended tests for association in 2×2 tables," *Statist. Med.*, 28, 1159–1175. MR2662203

Mehta, C. R., Patel, N. R., and Gray, R. (1985), "Computing an exact confidence interval for the common odds ratio in several 2×2 contingency tables," *J. Am. Statist. Assoc.*, 80, 969–973. MR0819602

Nissen, S. E. and Wolski, K. (2007), "Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes," *N. Engl. J. Med.*, 356, 2457–2471.

Sweeting, M. J., Sutton, A. J., and Lambert, P. C. (2004), "What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data," *Statist. Med.*, 23, 1351–1375.

Tian, L., Cai, T., Pfeffer, M. A., Piankov, N., Cremieux, P. Y., and Wei, L. J. (2009), "Exact and efficient inference procedure for meta-analysis and its application to the analysis of independent 2×2 tables with all available data but without artificial continuity correction," *Biostatistics*, 10, 275–281.

Vollset, S. E., Hirji, K. F., and Elashoff, R. M. (1991), "Fast computation of exact confidence limits for the common odds ratio in a series of 2×2 tables," *Journal of the American Statistical Association*, 86, 404–409.

Yang, G., Liu, D., Wang, J., and Xie, M. (2016), "Meta-analysis framework for exact inferences with application to the analysis of rare events," *Biometrics*, 72, 1378–1386.

Zhang, J., Carlin, B. P., Neaton, J. D., Soon, G. G., Nie, L., Kane, R., Virnig, B. A., and Chu, H. (2014), "Network meta-analysis of randomized clinical trials: Reporting the proper summaries," *Clinical Trials*, 11, 246–262.

Min-ge Xie
Professor
Department of Statistics and Biostatistics
Rutgers University
Piscataway, NJ 08854
USA
E-mail address: mxie@stat.rutgers.edu

John Kolassa
Professor
Department of Statistics and Biostatistics
Rutgers University
Piscataway, NJ 08854
USA
E-mail address: kolassa@stat.rutgers.edu

Dungang Liu
Assistant Professor of Business Analytics
Lindner College of Business
University of Cincinnati
Cincinnati, OH 45221
USA
E-mail address: dungang.liu@uc.edu

Regina Liu
Professor
Department of Statistics and Biostatistics
Rutgers University
Piscataway, NJ 08854
USA
E-mail address: rliu@stat.rutgers.edu

Sifan Liu
Postdoctoral Associate
Department of Statistics and Biostatistics
Rutgers University
Piscataway, NJ 08854
USA
E-mail address: sifan.liu@rutgers.edu