



# Something Out of Nothing? The Influence of Double-Zero Studies in Meta-analysis of Adverse Events in Clinical Trials

Zhaohu Fan<sup>1</sup> · Dungang Liu<sup>4</sup> · Yuejie Chen<sup>2</sup> · Nanhua Zhang<sup>3</sup>

Received: 24 March 2023 / Revised: 24 November 2023 / Accepted: 28 January 2024  
© The Author(s) 2024

## Abstract

In addition to clinical efficacy, safety is another important outcome to assess in randomized controlled trials. It focuses on the occurrence of adverse events, such as stroke, deaths, and other rare events. Because of the low or very low rates of observing adverse events, meta-analysis is often used to pooled together evidence from dozens or even hundreds of similar clinical trials to strengthen inference. A well-known issue in rare-event meta-analysis is that some or even majority of the available studies may observe zero events in both the treatment and control groups. The influence of these so-called double-zero studies has been researched in the literature, which nevertheless focuses on reaching a dichotomous conclusion—whether or not double-zero studies should be included in the analysis. It has not been addressed *when and how* they contribute to inference, especially for the odds ratio. This paper fills this gap using comparative analysis of real and simulated data sets. We find that a double-zero study contributes to the odds ratio inference through its sample sizes in the two arms. When a double-zero study has an unequal allocation of patients in its two arms, it may contain non-ignorable information. Exclusion of these studies, if taking a significant proportion of the study cohort, may result in inflated type I error, deteriorated testing power, and increased estimation bias.

**Keywords** Adverse event · Clinical trial · Odds ratio · Rare event · Unequal allocation · Zero event

## 1 Introduction

Efficacy and safety are pivotal issues in clinical treatments and procedures. Clear and sufficient evidence, demonstrating a new treatment or procedure is effective and safe, is required before its approval by the Food and Drug Administration [18]. Different from efficacy assessment aiming to prove effectiveness, a safety study seeks to identify adverse effects potentially associated with the new treatment [18]. The

---

Extended author information available on the last page of the article

adverse effects are detected by capturing certain adverse events, which are those events with dangerous and unwanted tendencies, such as death, stroke, and heart failure. Oftentimes, the adverse events of interest are rarely observed, as their occurrence rate could be extremely low. As a result, a single clinical trial, even with hundreds of patients enrolled, may be powerless to detect safety issues, if there is any. In this so-called rare-event situation, synthesis of the evidence from dozens or even hundreds of clinical trials may be the only way to make a reliable and meaningful conclusion [8]. This evidence synthesis process is often termed as meta-analysis in statistical and clinical literature. In the past decade, meta-analysis has been recognized and demonstrated as an effective tool in the discovery of safety issues. For example, in a high-profile meta-analysis of 48 clinical trials, Nissen and Wolski [12] concluded that the diabetic drug Rosiglitazone had a significant association with myocardial infarction. Such a discovery would not be possible if the 48 trials were not pooled together, as an analysis of any single-trial data did not yield statistical significance.

A common issue in meta-analysis of adverse events is the existence of the so-called double-zero studies. A double-zero study is a situation in which no adverse events occur in either the control or treatment arm of a study. Such studies may take a great proportion in the study cohort when the event rates are very low. There are extensive discussions on how to deal with double-zero studies in the literature (see, e.g., [2, 6, 8, 9, 19, 26]). A consensus so far is that the common continuity correction (i.e., adding 0.5 to zeros) may result in severe bias [8, 9, 14, 19]. However, there is still an unsettling question—when and how double-zero studies contribute to inference?

The impact of double-zero studies on statistical inference for the odds ratio (or relative risk) has been researched in the literature. Note that the odds ratio and relative risk have similar values when event rates are low. Some argue that in theory such studies provide information for inference. For example, Xie et al. [23] showed that double-zero studies can contribute to the full likelihood of the common odds ratio and therefore they contain information for meta-analysis inference. Nevertheless, others have reported numerical results that are not consistent or even contradictory to each other. For example, in the analysis of 60 clinical trials on the coronary artery bypass grafting, Kuss [7] observed the relative risk changed little with the 35 double-zero studies either included or excluded. In their development of an exact meta-analysis approach, Liu et al. [9] and Yang et al. [26] observed that the inclusion of double-zero studies always results in wider confidence intervals for the odds ratio and relative risk, which implies that the influence of double-zero studies may make inference conservative and potentially less efficient. Ren et al. [14] showed odds ratio disagreement between inclusion and exclusion of double-zero studies (c.f. Tables 3–4 therein) in 386 real data meta-analyses. But their comparison in this regard was limited to the inverse-variance method and Mantel-Haenszel method which required continuity corrections to incorporate double-zero studies. Xu et al. [25] used the generalized linear mixed model and performed a similar comparison in 442 meta-analyses. They reported noticeable numerical changes (e.g., odds ratio direction or statistical significance) when double-zero studies were excluded from meta-analysis. Xu et al. [24] used Doi's inverse variance heterogeneity (IVhet) model with continuity corrections (of 0.5) to include double-zero studies, which led to

improved performance. However, they noted that this may be due to the addition of 0.5 to zero cells. So far, the focus of the existing literature is on reaching a dichotomous conclusion—whether or not double-zero studies should be included in the analysis. There is still a lack of clear understanding of *when and how* double-zero studies contribute to the odds ratio inference.

The goal of this paper is to use real and simulated data analyses to explain *when and how* double-zero studies contribute to the odds ratio inference. We use as a prototype of the cohort of Kuss [7]’s 60 clinical trials on the coronary artery bypass grafting (CABG) to generate new data for our numerical investigation. We use a classical binomial-normal hierarchical model [4, 16] and conduct fixed- and random-effects analysis. Our finding is that a double-zero study contribute to the odds ratio inference through its sample sizes in the two arms. Roughly speaking, a double-zero study with an unequal allocation in the two arms (e.g.,  $n_c > n_t$ ) contains non-ignorable information for inference. Exclusion of these studies will lead to inflated type I error, deteriorated testing power, and increased estimation bias.

The rest of the article is organized as follows. In Sect. 2, we review the binomial-normal hierarchical model and a fully Bayesian inference for the odds ratio. In Sect. 3, we conduct a case study of Kuss [7]’s CABG data, through which we explain heuristically and numerically how the arm sizes of a double-zero study contribute to inference. The insights are used to guide the design of simulation studies in Sect. 4, where the impact of double-zero studies is demonstrated. In Sect. 5, we repeat the same investigation but in a random-effects model setting. The paper is concluded with a discussion in Sect. 6.

## 2 A Classical Meta-analysis Model for Odds Ratios

Given  $K$  independent clinical trials, we use the classical binomial-normal hierarchical model to make inference [4, 16, 17]. We assume that in the  $i$ th study, the numbers of (adverse) events  $Y_{ci}$  and  $Y_{ti}$ , in the control and treatment arms, respectively, follow binomial distributions

$$Y_{ci} \sim \text{Binomial}(n_{ci}, p_{ci}), \quad Y_{ti} \sim \text{Binomial}(n_{ti}, p_{ti}), \quad i = 1, \dots, K,$$

where  $n_{ci}$  and  $n_{ti}$  are the numbers of participants. The goal is to compare the probability  $p_{ci}$  of the control arm with the probability  $p_{ti}$  of the treatment arm. To gauge the difference between  $p_{ci}$  and  $p_{ti}$ , we consider the odds ratios  $\theta_i = \frac{p_{ti}}{1-p_{ti}} / \frac{p_{ci}}{1-p_{ci}}$ . The odds ratio is a risk measure commonly used in clinical trials, and it is studied in [9, 14, 23] in meta-analysis of rare events.

The logarithm of the odds ratios is  $\delta_i = \log(\theta_i) = \log\left(\frac{p_{ti}}{1-p_{ti}}\right) - \log\left(\frac{p_{ci}}{1-p_{ci}}\right) = \text{logit}(p_{ti}) - \text{logit}(p_{ci})$ . Using the log odds ratios  $\delta_i$ , the model can be reparameterized as

$$\text{logit}(p_{ci}) = \mu_{ci}, \quad \text{logit}(p_{ti}) = \mu_{ci} + \delta_i, \tag{1}$$

where  $\mu_{ci}$  is the baseline probability in the control arm of the  $i$ th study. In the classical fixed-effects model, the treatment effects  $\delta_i$  are fixed and identical across all

the studies, i.e.,  $\delta_i = \delta$ . The fixed-effects assumption can be relaxed to allow  $\delta_i$  to vary across studies. This so-called random-effects model will be examined later in this paper. In both fixed- and random-effects models, the baseline effects  $\mu_{ci}$  are allowed to vary between studies. Oftentimes, they are assumed to follow a normal distribution

$$\mu_{ci} \sim \mathcal{N}(a, b^2), \tag{2}$$

where  $a$  and  $b$  are nuisance parameters.

In this paper, we use a fully Bayesian method to make inference about  $\delta$ . We follow the convention to specify its non-informative prior as  $\delta \sim \mathcal{N}(0, 10^4)$ . The non-informative priors for the two nuisance parameters  $a$  and  $b$  in Eq. (2) can be specified as  $a \sim \mathcal{N}(0, 10^4)$  and  $b^2 \sim \mathcal{IG}(10^{-3}, 10^{-3})$  (see, e.g., [16, 17, 21]). We draw posterior samples using Markov Chain Monte Carlo via Gibbs sampling. Our implementation of Gibbs sampling uses the R package `Rjags` which calls the computing programs in JAGS [13]. Specifically, we run three Markov chains with distinct starting values to ensure that they converge to the same distribution. In each Markov chain, we use 10,000 burn-in iterations followed by 50,000 iterations to collect posterior samples. To reduce autocorrelation within the samples, we implement a thinning process, selecting every fourth value from the posterior samples of each chain.

### 3 A Case Study of Coronary Artery Bypass Grafting

Ischaemic heart disease refers to a condition of insufficient blood supply to the myocardium. A medical therapy for this condition is coronary artery bypass grafting (CABG) surgery (see, e.g., [1, 10, 27]). Traditionally, a CABG surgery is performed with cardiopulmonary bypass to provide artificial circulation, and the coronary artery bypass can be performed with the heart stopped. This procedure is called “on-pump” CABG. The on-pump CABG operation, however, may result in adverse events, such as myocardial, pulmonary, renal, coagulation, and cerebral complications (see, e.g., [11, 20]). In an effort to reduce the occurrence of adverse events, “off-pump” CABG, a relatively new procedure that does not require cardiopulmonary bypass, has been developed and used in recent years (see, e.g., [5, 11, 15]).

To compare the off-pump and on-pump methods, [11] collected 60 studies to examine the occurrence of postoperative strokes. Out of the 60 studies, 35 studies did not observe any postoperative strokes in either arm. The full data set is displayed in Table 1. In their analysis, [11] calculated the relative risk using the standard inverse-variance method. The relative risk was 0.73 with a 95% confidence interval of [0.53, 0.99] with the  $p$  value being 0.04. Therefore, they concluded that the off-pump method results in lower chance of postoperative strokes. Yet, their analysis ignored all of the 35 double-zero studies, which accounted for 58.3% of the available studies.

To investigate the impact of the 35 double-zero studies, [7] compared the results by excluding and including them in the beta-binomial model (which is different from

**Table 1** Data of the 60 independent studies comparing the off-pump and on-pump methods used in the coronary artery bypass grafting surgery with regard to the occurrence of postoperative strokes [11]

Study	Off-pump		On-pump	
	Strokes	No. of patients	Strokes	No. of patients
OCTOPUS2001	2	142	5	139
BHACAS I+II 2002	3	200	6	201
SMART 2003	2	100	2	100
AI-Ruzzeh 2006	2	84	1	84
DOORS 2009	10	450	18	450
MASS III 2009	3	156	5	155
ROOBY 2009	14	1104	8	1099
PROMISS 2010	0	73	0	74
BBS 2011	16	176	11	163
Matata 2000	0	10	0	10
Penttila 2001	0	11	0	11
Caputo 2002	0	20	0	20
Zamvar 2002	0	30	0	30
Carrier 2003	0	32	1	33
Raja 2003	3	150	4	150
Gerola 2004	0	80	0	80
PRAGUE-4 2004	0	208	2	192
Legare 2004	2	150	0	150
Lingaas 2004	0	60	2	60
Gasz 2005	0	10	0	20
JOCRI 2005	0	81	1	86
Ascione 2005	0	10	0	10
Niranjan 2006	1	40	1	40
Michaux 2006	0	25	0	25
Ascione 2006	0	20	0	20
Motallebzadeh 2006	1	108	5	104
Tatoulis 2006	0	50	0	50
Ozkara 2007	0	22	0	22
Hernandez 2007	0	102	3	102
Rasmussen 2007	0	18	0	17
Mandak 2008	0	20	0	20
Vural 1995	0	25	0	25
Gulielmos 1999	0	20	0	20
Czerny 2000	0	15	0	15
Diegeler 2000	0	20	1	20
Kochamba 2000	0	29	0	29
Wandschneider 2000	0	52	0	67
Czerny 2001	0	40	0	40
Sahlman 2003	1	24	1	26
Muneretto 2003	0	88	2	88

**Table 1** (continued)

Study	Off-pump		On-pump	
	Strokes	No. of patients	Strokes	No. of patients
Lee 2003	0	30	1	30
Vedin 2003	0	33	0	37
Velissaris 2003	0	27	0	27
Parolari 2003	0	11	0	14
Motallebzadeh 2004	0	15	1	20
Selvanayagam 2004	0	30	1	30
Gasz 2004	0	10	0	10
Synnergren 2004	0	26	0	26
Blacher 2005	0	13	0	15
Rachwalik 2006	0	21	0	21
Malik 2006	0	25	0	25
Cavalca 2006	0	25	0	25
Gnenc 2006	0	30	0	12
Rainio 2007	0	10	0	10
Kunes 2007	0	17	0	17
Parolari 2007	0	14	0	15
Sajja 2007	0	60	1	60
Jares 2007	0	10	1	10
Formica 2009	0	30	0	30
Modine 2010	0	35	0	36

the model used in [11]). When the double-zero studies were excluded, the relative risk was 0.53 with a 95% confidence interval of [0.31,0.91]. When the double-zero studies were included, the relative risk was 0.51 with a 95% confidence interval of [0.28, 0.92]. The results in the two scenarios are similar to each other, which suggests that information in the double-zero studies may be negligible for inference. However, this conclusion contradicts arguments made in other existing publications [9, 19, 22, 24–26].

Our intuition is that a double-zero study may contain useful information in the sample size of each of its arm. To illustrate this, we start with a toy example and then use the proof-by-contradiction method to re-analyze the CABG data set. Suppose the number of adverse events follows a binomial distribution  $Y \sim \text{Binomial}(n, p)$ , and we observe no event, i.e.,  $y = 0$ . If the sample size  $n = 10$ , the Bayesian method with a non-informative prior  $\text{Beta}(1, 1)$  on  $p$  yields a mean estimate  $\hat{p} = 0.083$  with a 95% credible interval (0.002, 0.285). However, if  $n = 1000$ , the same inference procedure yields a mean estimate  $\hat{p} = 0.001$  with a 95% credible interval (0.000, 0.004). The upper end 0.004 is much smaller than 0.285, and it is closer to 0. This implies that observing a zero event out of a larger sample size gives more confidence that the underlying probability is closer to zero. In other words, the same observation  $y = 0$  but with different sample sizes ( $n = 10$  or  $n = 1000$ ) may yield very different results.

To further manifest how the sample sizes of the two arms of a double-zero study may contribute to meta-analytical inference, we use the proof-by-contradiction method. We begin with an assumption that a double-zero study does *not* contribute any information to statistical inference, even through the sample sizes of its two arms. If this assumption is true, we can arbitrarily change its  $n_{ci}$  or  $n_{ti}$  and expect no or minimal numerical change in analysis result. For the CABG data set, we increase  $n_{ci}$  of the control arms in the 35 double-zero studies (only) and monitor the change in numerical results. Specifically, we multiply  $n_{ci}$  in the double-zero studies by the factors 2, 3, 4, and 5. Note that we do not alter the sample sizes in non-double-zero studies. We only alter the control-arm size  $n_{ci}$  in the 35 double-zero studies, which means any numerical change observed in subsequent analysis is due to the alteration of  $n_{ci}$  in the context of observing 0 event out of  $n_{ci}$  patients in the control arm.

Table 2 shows the odds ratio estimates and credible intervals obtained using the method presented in Sect. 2. When the scale factor is 1, results are shown for the analysis of the original CABG data set (the first row of Table 2). When the scale factor is larger than 1, the sample size  $n_{ci}$  of the control arm of the 35 double-zero studies has been increased. Table 2 shows that when all the 35 double-zero studies are excluded from the analysis, the odds ratio estimates and credible intervals change little, regardless of the scale factors. This is in line with our expectation as we have only altered  $n_{ci}$  in the double-zero studies and the data in those non-double-zero studies remain the same. On the other hand, when the 35 double-zero studies are included in the analysis, scaling up  $n_{ci}$  clearly moves the odds ratio estimate toward 1 (i.e., from 0.705 to 0.787). In particular, when the scaling factor is 2 (e.g.,  $n_{ci}$  is doubled), the upper end of the credible interval becomes  $1.006 > 1$ , which indicates a *non-significant* difference between the off-pump and on-pump methods. This is in contrast with the analysis of the original data set (scale factor = 1) where the upper limit of the credible interval is  $0.942 < 1$ , which indicates a significant difference between the two surgical methods. To summarize, the numerical changes seen after  $n_{ci}$  of the double-zero studies scaled up contradicts the assumption that a double-zero study does not contribute any information to inference. In fact, our experiment here (by altering  $n_{ci}$  in double-zero studies only) has provided numerical evidence that  $n_{ci}$  (or  $n_{ti}$ ) can contribute to meta-analysis ignorably information that may change the significance conclusion.

**Table 2** Analysis of the CABG data set in [11] with  $n_{ci}$  artificially increased

Scale factor	Excluding double-zero studies		Including double-zero studies	
	Odds ratio estimate	95% Credible interval	Odds ratio estimate	95% Credible interval
1	0.705	[0.473, 0.942]	0.705	[0.475, 0.942]
2	0.706	[0.468, 0.945]	0.756	[0.499, 1.006]
3	0.708	[0.477, 0.949]	0.779	[0.519, 1.044]
4	0.705	[0.472, 0.946]	0.785	[0.517, 1.051]
5	0.705	[0.477, 0.945]	0.787	[0.528, 1.063]

The numerical evidence in this section has provides critical insights into when and how a double-zero study may contribute to the inference. Roughly put, if a double-zero study contributes to the overall inference, it is through the sample sizes  $n_{ci}$  and  $n_{ti}$  of its two arms, and this contribution is more manifest when its two arms have unequal allocations (e.g.,  $n_{ci} \gg n_{ti}$ ). The simulation studies in the following section will further demonstrate this heuristic statement by considering different allocations in the two arms in trial designs.

## 4 Simulation studies

The goal of our simulation studies is to compare the “full analysis” that includes all available data with “partial analysis” that excludes double-zero studies. We assess their performance with respect to the type I error, testing power, and estimation bias. We consider a variety of settings by varying (1) the odds ratio, (2) the number of subjects in each clinical study, and (3) the total number of clinical studies. Our results are based on 1000 simulation repetitions.

We simulate  $Y_{ci}$  and  $Y_{ti}$  for  $K$  independent studies from Model (1). To ensure low or very low baseline event rates, we set the parameters Eq. (2) as  $a = \text{logit}(p_{\max}/2)$  and  $b = (\text{logit}(p_{\max}) - a)/3$ . The value of  $p_{\max}$  controls the upper bound of the baseline probabilities  $p_{ci}$ . When  $p_{\max} = 1\%$  or  $0.5\%$ , we summarize the distribution of  $p_{ci}$ ’s in Table 3. For example, when  $p_{\max} = 1\%$ , the 99% quantile of the the baseline probabilities is  $0.86\%$ .

### 4.1 Equal allocation

We begin with an equal allocation setting where the sample sizes of the treatment and control groups are the same for each clinical study (e.g.,  $n_{ci} = n_{ti} = 200$ ). We set (i) the total number of clinical studies  $K = 60$  or  $180$ , (ii)  $p_{\max} = 0.5\%$ , and (iii) the odds ratio  $\theta = 1, 1.2, 1.4, 1.6, 1.8$  and  $2$ . Table 4 shows that the varying odds ratios result in different percentages of the double-zero studies ranging between  $23\%$  and  $37\%$ . Our goal is to test testing power when the null hypothesis is  $H_0$ : the odds ratio  $\theta = 1$ .

We compare the full and partial analyses in terms of the type I error and testing power. When  $K = 60$ , the top panel of Table 4 shows that these two types of analyses produce similar type I error rates, both of which are close to the  $5\%$  nominal level. The partial analysis also yields similar testing power to that of the full analysis across all the odds ratios greater than  $1$ . This observation implies that in this equal allocation setting, the partial analysis may perform as well as the full analysis. In other words, the inclusion/exclusion of double-zero studies has little influence on the inference

**Table 3** Distribution of the baseline probabilities  $p_{ci}$

$p_{\max}$	Mean	Standard deviation	95% quantile	99% quantile
1.00%	0.51%	0.12%	0.73%	0.86%
0.50%	0.26%	0.06%	0.37%	0.43%



**Table 4** Power for testing the null hypothesis  $H_0 : \theta = 1$  (equal allocation)

Odds ratio	1.0	1.2	1.4	1.6	1.8	2.0
<i>K</i> = 60						
Full analysis	5.30%	6.00%	16.90%	38.40%	59.80%	80.70%
Partial analysis w/o double-zero studies	5.30%	6.30%	16.90%	37.40%	59.70%	81.70%
Average number of double-zero studies	22	20	18	16	15	14
<i>K</i> =180						
Full analysis	4.60%	17.50%	61.80%	91.60%	99.10%	99.90%
Partial analysis w/o double-zero studies	4.60%	17.80%	61.10%	91.60%	99.40%	99.90%
Average number of double-zero studies	67	61	55	50	45	41

even when double-zero studies take a significant proportion (23–37%) of the available studies. This conclusion holds when the number of studies increases to  $K = 180$  as seen in the bottom panel of Table 4.

In Table 5, we examine the bias in odds ratio estimates produced by the partial and full analyses. The results show that both partial and full analyses yield small biases, which are comparable to each other. This observation suggests that in this equal allocation setting, the inclusion/exclusion of double-zero studies has negligible influence on the estimation bias. This conclusion holds for both  $K = 60$  and  $K = 180$ .

**Remark 1** We use Gelman–Rubin diagnostic test to assess convergence of Markov chains [3]. For example, Table 6 presents a summary of the Gelman–Rubin diagnostic test statistics for the equal allocation setting where  $K = 60$ . The mean and maximum statistics show that Gelman–Rubin diagnostic test statistics in all 1000 simulation replications are close to 1, indicating satisfactory convergence of Markov Chains.

### 4.2 Unequal allocation

*Unequal allocation* ( $n_{ci} > n_{ti}$ ). We continue to carry out our comparative analyses but using an unequal allocation setting, where the sample sizes of the treatment and control groups could be very different. The simulation setting is similar to the previous one except that 120 out of 180 clinical studies are unequally allocated. Specifically, we set  $n_{ci} = 100$  in the control arm and  $n_{ti} = 50$  in the treatment arm. For the remaining 60 studies, the treatment and control groups have equal sample sizes (e.g.,  $n_{ci} = n_{ti} = 200$ ). The total sample size ( $n_{ci} + n_{ti}$ ) is not comparable across all the studies, which is intentional as it is not uncommon (see the meta-data in [12]).

In this setting that contains unequal-allocation studies, we continue to compare the full and partial analyses in terms of the type I error and testing power. When

**Table 5** Bias of the odds ratio estimates (equal allocation)

Odds ratio	1.0		1.2		1.4		1.6		1.8		2.0	
	Mean (se)	Bias	Mean (se)	Bias	Mean (se)	Bias	Mean (se)	Bias	Mean (se)	Bias	Mean (se)	Bias
<i>K</i> = 60												
Full analysis	1.072 (0.302)	0.072	1.289 (0.349)	0.089	1.510(0.396)	0.110	1.725 (0.443)	0.125	1.935 (0.487)	0.135	2.146 (0.537)	0.146
Partial analysis w/o double-zero studies	1.074 (0.304)	0.074	1.289 (0.350)	0.089	1.512 (0.397)	0.112	1.726 (0.444)	0.126	1.939 (0.489)	0.139	2.147 (0.538)	0.147
<i>K</i> = 180												
Full analysis	1.021 (0.150)	0.021	1.226 (0.176)	0.026	1.431(0.200)	0.031	1.635 (0.220)	0.035	1.839 (0.238)	0.039	2.042 (0.261)	0.042
Partial analysis w/o double-zero studies	1.021 (0.151)	0.021	1.226 (0.177)	0.026	1.431 (0.200)	0.031	1.636 (0.221)	0.036	1.839 (0.240)	0.039	2.043 (0.262)	0.043

**Table 6** Summary of Gelman–Rubin diagnostic test statistics in 1000 simulation replications for the equal allocation setting where  $K = 60$

Odds ratio	1.0	1.2	1.4	1.6	1.8	2.0
Full analysis						
Mean	1.002	1.001	1.002	1.002	1.002	1.002
Maximum	1.018	1.017	1.018	1.017	1.017	1.018
Minimum	1.000	1.000	1.000	1.000	1.000	1.000
Partial analysis						
Mean	1.003	1.003	1.002	1.003	1.003	1.003
Maximum	1.020	1.019	1.020	1.020	1.019	1.020
Minimum	1.000	1.000	1.000	1.000	1.000	1.000

$p_{\max} = 0.5\%$ , the number of double-zero studies in each meta-analysis ranges from 86 to 104 (the proportion ranging from 48 to 58%). The top panel of Table 7 shows that the type I error produced by the partial analysis is 10.10%, which is twice as large as the 5% nominal level. In contrast, the full analysis yields a type I error of 4.80%, which is close to the 5% nominal level. Furthermore, the partial analysis has much lower testing power across all the odds ratios greater than 1, when compared to that of the full analysis. For example, if we examine the odds ratio = 1.4 in the third column of Table 7, the testing power of the partial analysis is 15.20%, which is less than a half of that of the full analysis (33.50%). These observations demonstrate that in this unequal allocation setting, the full analysis outperforms the partial analysis. The exclusion of double-zero studies from the analysis can result in substantially inflated type I error and substantially undermined testing power. When  $p_{\max}$  increases to 1%, the bottom panel of Table 7 shows that the exclusion of double-zero studies severely inflates the type I error (e.g., 11.30%) and adversely undermines its power as well (e.g., 10.30% as compared to 21.90% from the full analysis when the odds ratio is 1.2).

In Table 8, we examine the bias in odds ratio estimates produced by the partial and full analyses. When  $p_{\max}=0.5\%$ , the top panel shows that the estimation bias in the partial analysis is double or even triple of those produced in the full analysis across all the odds ratios considered. For example, when the odds ratio

**Table 7** Power for testing the null hypothesis  $H_0 : \theta = 1$  (unequal allocation:  $n_{ci} > n_{it}$ )

Odds ratio	1.0	1.2	1.4	1.6	1.8	2.0
$p_{\max} = 0.5\%$						
Full analysis	4.80%	9.80%	33.50%	66.20%	88.20%	97.20%
Partial analysis w/o double-zero studies	10.10%	4.70%	15.20%	43.80%	73.90%	90.90%
Average number of double-zero studies	104	100	97	93	90	86
$p_{\max}=1\%$						
Full analysis	6.00%	21.90%	67.60%	95.00%	99.60%	100.00%
Partial analysis w/o double-zero studies	11.30%	10.30%	46.90%	86.40%	97.80%	99.99%
Average number of double-zero studies	65	61	57	54	50	48

**Table 8** Bias of the odds ratio estimates

Odds Ratio	1.0		1.2		1.4		1.6		1.8		2.0	
	Mean (se)	Bias	Mean (se)	Bias	Mean (se)	Bias	Mean (se)	Bias	Mean (se)	Bias	Mean (se)	Bias
$p_{\max} = 0.5\%$												
Full analysis	1.03(0.20)	0.03	1.23 (0.23)	0.03	1.44(0.26)	0.04	1.64 (0.29)	0.04	1.85 (0.32)	0.05	2.05 (0.35)	0.05
Partial analysis w/o double- zero studies	0.93(0.17)	-0.07	1.11 (0.21)	-0.09	1.30(0.23)	-0.10	1.48 (0.26)	-0.12	1.67 (0.29)	-0.13	1.86(0.31)	-0.14
$p_{\max} = 1\%$												
Full analysis	1.02(0.15)	0.02	1.22 (0.17)	0.02	1.42(0.19)	0.02	1.63 (0.21)	0.03	1.83 (0.23)	0.03	2.03 (0.25)	0.03
Partial analysis w/o double- zero studies	0.94(0.14)	-0.06	1.14 (0.16)	-0.06	1.33(0.18)	-0.07	1.52 (0.20)	-0.08	1.72 (0.22)	-0.08	1.91(0.24)	-0.09

is 1.2, the bias produced from the partial analysis is  $-0.09$ , which is three times as large as bias (0.03) produced by the partial analysis. When  $p_{\max} = 1\%$ , our observation in the bottom panel of Table 8 is similar. The full analysis shows much smaller bias. These observations again confirms double-zero studies can contain useful information for inference in the unequal-allocation setting, and including them in the analysis can decrease the estimation bias.

**Remark 2** We carry out additional simulations when the total number of studies  $K$  is much smaller (e.g.,  $K = 60$ ), and the total sample size ( $n_{ci} + n_{ti}$ ) is comparable across all the clinical studies (e.g., ( $n_{ci} = 300, n_{ti} = 100$ ) and ( $n_{ci} = 200, n_{ti} = 200$ )). The results are similar and can be found in Supplementary Materials A.

*Unequal allocation ( $n_{ci} < n_{ti}$ ).* We consider another unequal-allocation setting but set  $n_{ci} = 200 < n_{ti} = 400$  in 120 studies, out of the 180 simulated studies. Treatment and control arms have equal sample sizes in the remaining 60 studies (i.e.,  $n_{ci} = n_{ti} = 200$ ).

Table 9 shows that the type I error produced by the partial analysis is slightly inflated (7.00%), whereas the type I error (6.20%) yielded by the full analysis is closer to the 5% nominal level. On the other hand, the full analysis yields higher testing power than that of the partial analysis across all the odds ratios greater than 1 (e.g., 49.60% versus 41.10% when the odds ratio = 1.3). For example, the full analysis has an increase of 2.0–8.5% in testing power across odds ratios, when compared to that of partial analysis. These observations indicate that the full analysis outperforms the partial analysis. Meta-analyses with double-zero studies can have greater testing power in an unequal-allocation setting.

In Supplementary Materials B, we compare bias in odds ratio estimates between partial and full analyses. The full analysis shows much smaller bias (see Table S3 for details).

**Table 9** Power for testing the null hypothesis  $H_0 : \theta = 1$  (unequal allocation:  $n_{ci} < n_{ti}$ )

Odds ratio	1.0	1.1	1.2	1.3	1.4	1.5	1.6
$p_{\max} = 0.5\%$							
Full analysis	6.20%	9.00%	23.00%	49.60%	74.00%	89.70%	96.40%
Partial analysis w/o double studies	7.00%	6.10%	17.90%	41.10%	66.40%	84.70%	94.40%
Average number of double-zero studies	49	46	43	40	37	34	32

**Table 10** Power for testing the null hypothesis  $H_0 : \theta = 1$  in the random-effects model

Odds ratio	1.0	1.2	1.4	1.6	1.8	2.0
Full analysis	6.90%	7.80%	30.20%	60.20%	83.80%	93.90%
Partial analysis w/o double-zero studies	11.30%	6.10%	21.30%	48.80%	78.10%	91.20%
Average number of double-zero studies	104	100	96	92	89	86

**Table 11** Random-effects analysis of the CABG data set in [11] with the sample size  $n_{ci}$  in the control arms of the double-zero studies artificially increased

Scale factor	Excluding double-zero studies			Including double-zero studies		
	Odds ratio est.	95% Credible interval	$\sigma$ est.	Odds ratio est.	95% Credible interval	$\sigma$ est.
1	0.599	[0.294, 0.893]	0.110	0.555	[0.230, 0.866]	0.123
2	0.590	[0.294, 0.882]	0.105	0.654	[0.305, 0.979]	0.103
3	0.602	[0.305, 0.899]	0.113	0.691	[0.356, 1.028]	0.086
4	0.598	[0.287, 0.901]	0.109	0.715	[0.396, 1.033]	0.093
5	0.598	[0.300, 0.889]	0.106	0.706	[0.361, 1.047]	0.098

### 5 Results in a Random-Effects Model Setting

In this section, we examine the difference between the partial and full analyses in a random-effects model setting. In the random-effects model, the treatment effects  $\delta_i$ 's are assumed to be drawn from a normal distribution  $\delta_i \sim \mathcal{N}(\delta, \sigma^2)$ . More specifically,

$$\begin{aligned} \text{logit}(p_{ci}) &= \mu_{ci}, \text{ logit}(p_{ti}) = \mu_{ci} + \delta_i, \quad i = 1, \dots, K. \\ \mu_{ci} &\sim \mathcal{N}(a, b^2), \delta_i \sim \mathcal{N}(\delta, \sigma^2). \end{aligned} \tag{3}$$

In the random-effects model, the parameter of interest is the mean parameter  $\delta$  of the distribution of  $\delta_i$  in Model (3). The non-informative priors for  $\delta$  and  $\sigma^2$  can be specified as  $\mathcal{N}(0, 10^4)$  and  $\mathcal{IG}(10^{-3}, 10^{-3})$ , respectively (see, e.g., [16, 21]). The priors for both  $a$  and  $b$  can be specified in the same way as in the fixed-effects model.

We simulate data from Model (3) to compare the full and partial analyses. We set the parameter  $\sigma=0.1$  to mimic the heterogeneity in the last column of Table 11. The rest of simulation settings is similar to that of the fixed-effects model in Sect. 4.2, **unequal allocation** ( $n_{ci} > n_{ti}$ ). Table 10 shows that the type I error produced by the partial analysis is severely inflated (11.30%), whereas the type I error (6.90%) yielded by the full analysis is close to the 5% nominal level. On the other hand, when the odds ratios are greater than 1, the full analysis yields higher testing power than the partial analysis (e.g., 60.20% versus 48.80% when the odds ratio=1.6). These

observations indicate the inclusion of double-zero studies increases testing power while reducing the type I error.

Similar to the real data analysis conducted in Sect. 3, we assume that a double-zero study does not contribute to statistical inference, even when taking into consideration the sample sizes of both arms. If this premise is true, we can alter the sample size  $n_{ci}$  or  $n_{ti}$  and expect little or no numerical change in the analysis results. Specifically, we increase the sample size of the control arm by multiplying it by the factors of 2, 3, 4, and 5.

Table 11 shows odds ratio estimates, credible intervals and estimates of the heterogeneity parameter  $\sigma$  obtained using the method presented in Sect. 5. When the scale factor is set to 1, analysis results are shown for the original CABG data set (the first row of Table 11). When the scale factor exceeds 1, the sample size  $n_{ci}$  of the control arm in the 35 double-zero studies has been increased. Table 11 shows when all of the 35 double-zero studies are removed from the analysis, the odds ratio estimates, credible intervals, and heterogeneity estimates remain almost unchanged, regardless of the scale factors. This is consistent with our expectations, as we have only modified  $n_{ci}$  in the double-zero studies, while the data in the non-double-zero studies remains unchanged (from the second to the fourth columns of Table 11). Conversely, when the 35 double-zero studies are incorporated into the analysis, increasing  $n_{ci}$  leads to a noticeable shift of the odds ratio estimate towards 1 (from 0.555 to 0.706). Furthermore, if the scaling factor is set to 3 (i.e.,  $n_{ci}$  is tripled), the upper end of the credible interval becomes  $1.028 > 1$ . This suggests that there is no significant difference between the off-pump and on-pump methods. This stands in contrast to the analysis of the original dataset (scale factor = 1), in which the upper bound of the credible interval is  $0.866 < 1$ , indicating a significant difference between the two surgical methods. In summary, the numerical alterations observed after scaling up  $n_{ci}$  in the double-zero studies contradict the assumption that a double-zero study does not provide any information for inference. In fact, our experiment here (by modifying  $n_{ci}$  solely in the double-zero studies) has provided numerical proof that  $n_{ci}$  (or  $n_{ti}$ ) can supply non-ignorable information to meta-analysis that may alter the conclusion of significance.

In Supplementary Materials B, we examine bias in odds ratio estimates produced by the partial and full analyses. The full analysis shows much smaller bias (see Table S4 for details).

## 6 Discussion

Rare adverse events, such as strokes and deaths, are of crucial concerns in safety studies of clinical treatments and procedures. However, their low occurrence rates pose challenges and questions for statistical inference. Many of them are surrounding double-zero studies. The debate concerning the inclusion and exclusion of double-zero studies was prominently sparked by the high-profile study by [12]. To examine the safety of the diabetic drug Rosiglitazone, they used Peto's method through which double-zero studies did not contribute anything to the inference. This practice has been questioned by many statisticians, leading to

two lines of research. One line has focused on the development of new methods that can incorporate double-zero studies without using 0.5 corrections (see the review article by [8]). The other line has attempted to reach a inclusion/exclusion conclusion, which is nevertheless dichotomous (see the references cited in the introduction). Before recommending an action (inclusion/exclusion), one question needs to be answered; that is, when and how double-zero studies contribute to the odds ratio inference. This is the main purpose of our investigation, and we intentionally avoid advocating a specific inclusion/exclusion action. As shown in our paper, double-zero studies may contribute significantly in some scenarios while in other scenarios they contribute little.

Through numerical studies, we have empirically found that when the following two conditions are met, double-zero studies may likely contribute to the odds ratio inference to a notable extent.

- (i) The group of non-double-zero studies itself contains adequate information for double-zero studies to borrow.
- (ii) There is a substantial number of double-zero studies with unequal allocation in the two arms.

Basically, to see notable impact of double-zero studies, Condition (i) says that the number of non-double-zero studies  $K_{\text{non-DZ}}$  can not be small. Otherwise, there is not adequate information to borrow from. Condition (ii) says that the number of double-zero studies with unequal allocations  $K_{\text{DZ-unequal}}$  should be (moderately) large. Otherwise, the impact of double-zero studies may not accumulate sufficiently to make a practical difference. These two conditions can be explained more intuitively in a non-meta-analysis setting where all the  $K$  studies have the same baseline probability and odds ratio. In this setting, we can stack the data throughout the  $K$  studies and form a single 2 by 2 contingency table to draw an inference. If Condition (i) is not met and  $K_{\text{non-DZ}}$  is too small, the odds ratio inference may be too unreliable as the numerators of the control and treatment arms may be small. If Condition (ii) is not met and  $K_{\text{DZ-unequal}}$  is not large, we may not see disproportionate increases in the denominators of the control and treatment arms and thus a difference in the odds ratio inference. An implication is that if the total number of studies  $K$  is small, at least one of the two conditions cannot be met, and therefore, the impact of double-zero studies may not be significant enough to make a clinical difference.

When the impact of double-zero studies is non-ignorable, our finding is that they contribute to odds ratio inference through the sample sizes in their two arms. When a double-zero study has unequal allocation in its two arms, it may contain non-ignorable information. If double-zero studies as such make up a large proportion of the study cohort, they should not be excluded from analysis. If excluded, the inference may not be valid or efficient as we have seen severely inflated type I error, deteriorated testing power, and increased estimation bias in our numerical study.

In practice, study level characteristics may explain the variation in the prevalences of adverse event outcomes as well as the differences between active treatment and control. These factors may be accounted for in meta-analyses through meta-regression. It is interesting to examine how double-zero studies may impact



inference especially when the site-level characteristics may be related to the occurrence of double-zero studies.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s12561-024-09431-y>.

## Declarations

**Ethical standards** No external funding or grants were received to support this work. Yuejie Chen was an undergraduate student at the University of Cincinnati and graduate student at North Carolina State University, and she is currently Data Analyst at LexisNexis, Raleigh, North Carolina. The real data presented in Table 1 were originally published in [11].

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Alexander JH, Smith PK (2016) Coronary-artery bypass grafting. *N Engl J Med* 374:1954–1964
- Bradburn MJ, Deeks JJ, Berlin JA, Russell Localio A (2007) Much ado about nothing: a comparison of the performance of meta-analytical methods with rare events. *Stat Med* 26:53–77
- Brooks SP, Gelman A (1998) General methods for monitoring convergence of iterative simulations. *J Comput Graph Stat* 7:434–455
- Günhan BK, Röver C, Friede T (2020) Random-effects meta-analysis of few studies involving rare events. *Res Synth Methods* 11:74–90
- Hannan EL, Wu C, Smith CR, Higgins RS, Carlson RE, Culliford AT, Gold JP, Jones RH (2007) Off-pump versus on-pump coronary artery bypass graft surgery: differences in short-term outcomes and in long-term mortality and need for subsequent revascularization. *Circulation* 116:1145–1152
- Sweeting JM, Sutton JA, Lambert PC (2004) What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. *Stat Med* 23:1351–1375
- Kuss O (2015) Statistical methods for meta-analyses including information from studies without any events—add nothing to nothing and succeed nevertheless. *Stat Med* 34:1097–1116
- Liu D (2019) Meta-analysis of rare events. In: Wiley StatsRef: statistics reference. Online 1–7. <https://doi.org/10.1002/9781118445112.stat08167>
- Liu D, Liu RY, Xie M-G (2014) Exact meta-analysis approach for discrete data and its application to 2 × 2 tables with rare events. *J Am Stat Assoc* 109:1450–1465
- Lorenzen US, Buggeskov KB, Nielsen EE, Sethi NJ, Carranza CL, Gluud C, Jakobsen JC (2019) Coronary artery bypass surgery plus medical therapy versus medical therapy alone for ischaemic heart disease: a protocol for a systematic review with meta-analysis and trial sequential analysis. *Syst Rev* 8:1–14
- Møller CH, Penninga L, Wetterslev J, Steinbrüchel DA, Gluud C (2012) Off-pump versus on-pump coronary artery bypass grafting for ischaemic heart disease. *Cochrane Database Syst Rev* 3:CD007224
- Nissen SE, Wolski K (2007) Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *N Engl J Med* 356:2457–2471
- Plummer M, Stukalov A, Denwood M (2022) rjags: Bayesian graphical models using MCMC. R package version 4.13

14. Ren Y, Lin L, Lian Q, Zou H, Chu H (2019) Real-world performance of meta-analysis methods for double-zero-event studies with dichotomous outcomes using the Cochrane Database of Systematic Reviews. *J Gen Intern Med* 34:960–968
15. Shroyer AL, Grover FL, Hattler B, Collins JF, McDonald GO, Kozora E, Lucke JC, Baltz JH, Novitzky D (2009) On-pump versus off-pump coronary-artery bypass surgery. *N Engl J Med* 361:1827–1837
16. Smith TC, Spiegelhalter DJ, Thomas A (1995) Bayesian approaches to random-effects meta-analysis: a comparative study. *Stat Med* 14:2685–2699
17. Sutton AJ, Abrams KR (2001) Bayesian methods in meta-analysis and evidence synthesis. *Stat Methods Med Res* 10:277–303
18. Thaul S (2012) How FDA approves drugs and regulates their safety and effectiveness. Congressional Research Service
19. Tian L, Cai T, Pfeffer MA, Piankov N, Cremieux P-Y, Wei L (2009) Exact and efficient inference procedure for meta-analysis and its application to the analysis of independent  $2 \times 2$  tables with all available data but without artificial continuity correction. *Biostatistics* 10:275–281
20. Van Dijk D, Jansen EW, Hijman R, Nierich AP, Diephuis JC, Moons KG, Lahpor JR, Borst C, Keizer AM, Nathoe HM et al (2002) Cognitive outcome after off-pump and on-pump coronary artery bypass graft surgery: a randomized trial. *J Am Med Assoc* 287:1405–1412
21. Warn DE, Thompson S, Spiegelhalter DJ (2002) Bayesian random effects meta-analysis of trials with binary outcomes: methods for the absolute risk difference and relative risk scales. *Stat Med* 21:1601–1623
22. Xiao M, Lin L, Hodges JS, Xu C, Chu H (2021) Double-zero-event studies matter: a re-evaluation of physical distancing, face masks, and eye protection for preventing person-to-person transmission of COVID-19 and its policy impact. *J Clin Epidemiol* 133:158–160
23. Xie M-G, Kolassa J, Liu D, Liu R, Liu S (2018) Does an observed zero-total-event study contain information for inference of odds ratio in meta-analysis? *Stat Interface* 11:327–337
24. Xu C, Furuya-Kanamori L, Islam N, Doi SA (2022) Should studies with no events in both arms be excluded in evidence synthesis? *Contemp Clin Trials* 122:e106962
25. Xu C, Li L, Lin L, Chu H, Thabane L, Zou K, Sun X (2020) Exclusion of studies with no events in both arms in meta-analysis impacted the conclusions. *J Clin Epidemiol* 123:91–99
26. Yang G, Liu D, Wang J, Xie M-G (2016) Meta-analysis framework for exact inferences with application to the analysis of rare events. *Biometrics* 72:1378–1386
27. Yusuf S, Zucker D, Passamani E, Peduzzi P, Takaro T, Fisher L, Kennedy J, Davis K, Killip T, Norris R et al (1994) Effect of coronary artery bypass graft surgery on survival: overview of 10-year results from randomised trials by the Coronary Artery Bypass Graft Surgery Trialists Collaboration. *Lancet* 344:563–570

## Authors and Affiliations

Zhaohu Fan<sup>1</sup> · Dungang Liu<sup>4</sup>  · Yuejie Chen<sup>2</sup> · Nanhua Zhang<sup>3</sup>

✉ Dungang Liu  
dungang.liu@uc.edu

Zhaohu Fan  
jonathan.fan@scheller.gatech.edu

Yuejie Chen  
yuejiechenyj@gmail.com

Nanhua Zhang  
nanhua.zhang@cchmc.org

<sup>1</sup> Information Technology Management, Scheller College of Business, Georgia Institute of Technology, Atlanta, GA, USA

<sup>2</sup> LexisNexis, Raleigh, NC, USA

<sup>3</sup> Division of Biostatistics and Epidemiology, Cincinnati Children's Hospital Medical Center and College of Medicine, University of Cincinnati, Cincinnati, OH, USA

<sup>4</sup> Department of Operations, Business Analytics and Information Systems, Carl H. Lindner College of Business, University of Cincinnati, Cincinnati, OH, USA