Original Article

Improvement of Statistical Power to Detect Publication Bias in Meta-analysis Using the Clinical Trial Registration System

Nobushige Matsuoka^{*1}, Hiroshi Horio^{*1} and Chikuma Hamada^{*2}

^{*1}Department of Management Science, Graduate School of Engineering,
Tokyo University of Science

^{*2}Department of Management Science, Faculty of Engineering,
Tokyo University of Science

e-mail:nobushige.matsuoka@pfizer.com

As clinical trials with “positive” results are more likely to be published, a meta-analysis of only published trials may be biased toward positive results (referred to as “publication bias”). A number of statistical tests have been proposed to detect publication bias. However, they have undesirable properties, particularly, the inflation of type I error and low power. A primordial countermeasure has been launched. In September 2004, the International Committee of Medical Journal Editors announced that they would no longer publish trials that were not registered in a public registry in advance. They embraced the WHO trial registration set consisting of 20 items including target sample size, which is related to the publication of results. The aim of this paper is to propose a new approach with a higher statistical power for detecting publication bias by using information on the sample sizes of all trials, including unpublished trials from the registry. We compared the proposed method to commonly used methods via simulations. The proposed method was found to have a higher power than the other methods in many situations. It will be useful for detecting publication bias because clinical trial registration will be more widespread in the near future.

Key words: meta-analysis; publication bias; clinical trial registration; Begg’s rank correlation method, Egger’s linear regression method, Macaskill’s linear regression method.

1. Introduction

Evidence-Based Medicine (EBM) is firmly radicated in the field of medicine. The meta-analysis (as part of a systematic review) of randomized controlled trials is considered the strongest evidence for this and it plays an important role in the realization of EBM. However, the reliability

of meta-analyses can be affected by various sources of bias, publication bias being the predominant one. In general, clinical trials with “statistically significant” or “positive” results are more likely to be published. Therefore, a meta-analysis of only published trials may be biased toward positive results, and the analysis of bias should be an important part of the meta-analysis.

A funnel plot (i.e., a scatter plot of the estimated treatment effect against a measure of its precision) is commonly used as a visual tool to detect publication bias. Usually, the total sample size or the inverse of the standard error is used as a measure of precision. While a symmetrical funnel-shaped plot indicates the absence of publication bias in meta-analysis, an asymmetrical one indicates the presence of publication bias in meta-analysis. However, the assessment of a funnel plot might be subjective, and a recent empirical evaluation suggests that the interpretations of funnel plot can be limited (Terrin et al., 2005).

A number of statistical tests have been proposed to detect publication bias. The most widely used methods are based on the association between the estimated treatment effect and its variance by using an adjusted rank correlation method (Begg and Mazumdar, 1994) or a linear regression method, in which the standard normal deviate of effect size is regressed against its precision (i.e., the inverse of standard error) (Egger et al., 1997). However, there have been considerable debates regarding the properties of such tests, in particular, the inflation of type I error and low power. An alternative test has been proposed on the basis of a weighted linear regression method in which the estimated treatment effect is regressed against total sample size (Macaskill et al., 2001). While this method does not suffer from an inflated type I error, its conservative nature reflects lower power.

A primordial countermeasure for publication bias has been launched. In September 2004, the International Committee of Medical Journal Editors (ICMJE) declared that they would require, as a condition of consideration for publication, registration in a public trial registry at or before the onset of patient enrollment (DeAngelis, 2004). The ICMJE embraced the WHO trial registration set consisting of 20 items including target sample size (WHO, 2009). Therefore, target sample sizes are available even if trials are not published, owing to the spread of clinical trial registration. The target sample sizes will provide helpful information to analyze the influence of publication bias, because several researchers have mentioned the relationship between sample size and publication status (Easterbrook et al., 1991; Song et al., 2000; Thornton and Lee, 2000; Rothstein et al., 2005).

The aim of this paper is to propose a new approach with a higher statistical power for detecting publication bias in meta-analysis by using information on sample sizes of unpublished trials. The notations and clinical trials assumed in this paper are described in Section 2. Three established methods for detecting publication bias are described in Section 3. The proposed method for detecting publication bias on the basis of the clinical trial registration system is presented in Section 4. The methods of generating the data and scenarios for the simulation

study are described in Section 5, and the simulation results are given in Section 6. A case study is described in Section 7. We give further consideration to be discussed in Section 8.

2. Notations and Clinical Trials Assumed

Let us assume that a meta-analysis of k randomized controlled trials with binary outcome data is conducted. For each studies of sample size n_i ($i = 1, \dots, k$), let a_i and c_i denote the number of patients with event in the treated and control groups, respectively, and b_i and d_i denote the number of patients without event in the treated and control group, respectively. We consider the log odds ratio θ as a measure of the treatment effect. Using observed frequencies, the estimated log odds ratio in the i th trial is given by $\hat{\theta}_i = \log \{(a_i d_i)/(b_i c_i)\}$, with the corresponding estimate of variance given by $s_i^2 = 1/a_i + 1/b_i + 1/c_i + 1/d_i$. In this paper, the observed events are supposed to be unfavorable clinical outcomes (e.g., death, clinical worsening). Therefore, a log odds ratio of < 0 (odds ratio < 1) represents a positive result.

As mentioned earlier, a funnel plot is commonly used as a visual tool to detect publication bias. Usually, the total sample size or the inverse of the standard error is used as a measure of precision. The estimated treatment effect with low precision will scatter widely at the bottom of the funnel plot. In the absence of publication bias, the plot resembles a symmetrical inverted funnel, as shown in Figure 1 (a). In the presence of publication bias, the funnel plot appears asymmetrical due to unpublished trials in the bottom right-hand side of the plot, as shown in Figure 1 (b).

Fig. 1. Example of funnel plots based on simulated meta-analyses (a) in the absence of publication bias and (b) in the presence of publication bias. The inverse of the standard error is used as the precision.

3. Established Methods to Detect Publication Bias

When conducting a meta-analysis, it should be verified that there is no publication bias. It means that a result of clinical trial is published independently of a result itself. The following

three established methods test the fundamental null hypothesis “there is no publication bias” by testing individual null hypothesis.

3.1 Begg’s rank correlation method

Begg and Mazumdar (1994) proposed an adjusted rank correlation method to evaluate the association between the standardized treatment effect $\hat{\theta}_i^*$ and its variance $v_i (= s_i^2)$ on the basis of Kendall’s rank correlation coefficient, where $\hat{\theta}_i^* = (\hat{\theta}_i - \hat{\theta}) / \sqrt{v_i^*}$, $\hat{\theta} = (\sum \hat{\theta}_i v_i^{-1}) / \sum v_i^{-1}$, and $v_i^* = v_i - (\sum v_j^{-1})^{-1}$.

The test is based on the derivation of Kendall’s rank correlation coefficient between $\hat{\theta}_i^*$ and v_i , which shows the association between the two quantities. Scatter plots of $\hat{\theta}_i^*$ and v_i based on simulated meta-analyses are shown in Figure 2. There is no correlation between $\hat{\theta}_i^*$ and v_i in the absence of publication bias (Figure 2 (a)). In contrast, there is negative correlation between $\hat{\theta}_i^*$ and v_i in the presence of publication bias (Figure 2 (b)).

Fig. 2. Example of Begg’s method based on simulated meta-analyses (a) in the absence of publication bias and (b) in the presence of publication bias.

Accordingly, we consider the following one-sided test at a 5% level of significance:

$$H_0: \tau = 0 \text{ v.s. } H_1: \tau < 0,$$

where τ is Kendall’s rank correlation coefficient.

3.2 Egger’s regression method

Egger et al. (1997) proposed an unweighted linear regression analysis to measure the asymmetry of a funnel plot. This method corresponds to a regression analysis of Galbraith’s radial plot (Galbraith, 1988). The standard normal deviate is regressed against the precision of estimate (1). This unweighted regression analysis can be shown to be identical to the weighted linear regression analysis given by (2), which corresponds to the funnel plot.

$$\left(\frac{\hat{\theta}_i}{s_i} \right) = \beta_0 + \beta_1 \left(\frac{1}{s_i} \right) + \epsilon_i \quad (1)$$

$$\hat{\theta}_i = \beta_1 + \beta_0 s_i + \epsilon_i, \text{ Weight: } 1/s_i^2 \quad (2)$$

In the absence of publication bias, the slope (β_0) in (2) has an expected value of zero, as shown in Figure 3 (a). However, in the presence of publication bias, β_0 has an expected value of < 0 in case the log odds ratio of < 0 represents a positive result, as shown in Figure 3 (b). Therefore, this regression test can be interpreted as a test of linear trend in a funnel plot (treatment effect against its precision).

Fig. 3. Example of Egger's method based on simulated meta-analyses (a) in the absence of publication bias and (b) in the presence of publication bias. The dotted lines are estimated regression lines.

Accordingly, we consider the following one-sided test at a 5% level of significance:

$$H_0: \beta_0 = 0 \text{ v.s. } H_1: \beta_0 < 0 \quad (3)$$

3.3 Macaskill's regression method

Macaskill et al. (2001) proposed a weighted linear regression analysis of the estimated treatment effect $\hat{\theta}_i$ on sample size n_i (4). In order to reduce the correlation between weight and estimated treatment effect, they suggested weighting by the reciprocal of the pooled variance for each trial, that is, the variance of the pooled estimate resulting from combining the data for two groups ($1/(a_i + c_i) + 1/(b_i + d_i)$).

$$\hat{\theta}_i = \beta_0 + \beta_1 n_i + \epsilon_i \quad (4)$$

In the absence of publication bias, the slope (β_1) in (4) has an expected value of zero, as shown in Figure 4 (a). However, in the presence of publication bias, β_1 has an expected value of > 0 , as shown in Figure 4 (b).

Accordingly, we consider the following one-sided test at a 5% level of significance:

$$H_0: \beta_1 = 0 \text{ v.s. } H_1: \beta_1 > 0 \quad (5)$$

While this test does not suffer from inflated type I error rates in the presence of a substantial treatment effect, its power is lower than that of Egger's regression method (Rothstein et al., 2005).

Fig. 4. Example of Macaskill's method based on simulated meta-analyses (a) in the absence of publication bias and (b) in the presence of publication bias. The dotted lines are estimated regression lines.

4. Improvement of Statistical Power Using the Clinical Trial Registration System

The ICMJE's aforementioned precondition for publication has been applied to all clinical trials starting enrollment after July 1, 2005. They embraced the WHO trial registration set consisting of 20 items (WHO, 2009). Of the 20 items, we take notice of the target sample size because several researchers have mentioned the relationship between sample size and publication status (Easterbrook et al., 1991; Song et al., 2000; Thornton and Lee, 2000; Rothstein et al., 2000). Information about the sample size of all registered trials can help detect publication bias. In fact, actual sample size is not always reached target sample size due to the difficulty of patient recruitment and so on. However, in this paper, we assumed that target sample size is achieved in all clinical trials, that is, actual sample size is equal to target sample size.

Let us assume that a total of m trials were registered prior to patient enrollment, of which k trials are published and l trials are unpublished ($m = k + l$). With the spread of clinical trial registration, the following dataset will be available to detect publication bias (Table 1).

Table 1. Dataset for the Detection of Publication Bias

Trial No.	log OR	Standard Error	(Target) Sample Size	Publication Status
1	$\hat{\theta}_1$	s_1	n_1	Published
2	$\hat{\theta}_2$	s_2	n_2	Published
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
k	$\hat{\theta}_k$	s_k	n_k	Published
$k + 1$	Unknown	Unknown	n_{k+1}	Unpublished
$k + 2$	Unknown	Unknown	n_{k+2}	Unpublished
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$k + l$	Unknown	Unknown	n_{k+l}	Unpublished

k : Number of published trials
 l : Number of unpublished trials

As mentioned in the previous section, the fundamental null hypothesis for detection of publication bias is that a result of trial is published independently of a result itself. In the proposed method, two different null hypotheses are considered in order to test this fundamental hypothesis. One is the same null hypothesis as the Egger's regression method (i.e., $\beta_0 = 0$). In addition to this, the null hypothesis regarding the sample sizes of published and unpublished trials is considered in the proposed method. If a clinical trial with "significant" or "positive" result is more likely to be published, sample size of unpublished trial tend to be small compared to published trial because a trial with small sample size may fail to demonstrate a significant treatment effect. Therefore, the null hypothesis "the distribution of the sample sizes of both published and unpublished trials are the same" is considered. To test these two hypotheses corresponds to testing the fundamental null hypothesis. Specifically, the following approach is conducted in the proposed method.

Two types of Z -scores are introduced in the proposed approach for the detection of publication bias. One (Z_E) represents the correlation between the estimated treatment effect and its precision on the basis of the Egger's regression method (i.e., Z_E is related to the test for the slope in (2)). The other (Z_W) is related to the comparison of sample sizes between published and unpublished trials on the basis of the Wilcoxon rank sum test. First, the Egger's regression method is applied to the data from published trials ($i = 1, \dots, k$) and the lower tailed p -value testing the null hypothesis " $H_0: \beta_0 = 0$ " is obtained. Then, Z_E corresponding to the p -value is calculated using the probit function $\Phi^{-1}(p)$.

Secondly, Z_W is calculated on the basis of the Wilcoxon rank sum test for comparison of the sample sizes between published ($i = 1, \dots, k$) and unpublished trials ($i = k + 1, \dots, k + l$). Let r_i denote the rank of each ordered n_i . Statistic W is defined as the sum of the ranks corresponding to the sample sizes of unpublished trials (6).

$$W = \sum_{i=k+1}^{k+l} r_i \quad (6)$$

The test statistic to test the null hypothesis " H_0 the distribution of the sample sizes of both published and unpublished trials are the same" is given as

$$Z_W = \frac{W - E_0(W)}{\sqrt{\text{Var}_0(W)}}, \quad (7)$$

where $E_0(W)$ and $\text{Var}_0(W)$ are the expectation and variance of W under the null hypothesis, respectively. For example, if there is no tie, $E_0(W) = \{l(k+l+1)\}/2$ and $\text{Var}_0(W) = \{kl(k+l+1)\}/12$.

Both Z_E and Z_W are (asymptotically) normally distributed with $N(0, 1^2)$ under each null hypothesis (i.e., there is no publication bias). Therefore, we introduce the following statistic Z as a test statistic for the detection of publication bias in the proposed method (8).

$$Z = \frac{Z_E + Z_W}{\sqrt{2}} \quad (8)$$

If Z_E and Z_W are independent and identically distributed, Z is normally distributed with $N(0, 1^2)$. The lower tailed p -value is given by $\Phi(Z)$, where $\Phi(\cdot)$ is the cumulative distribution function of $N(0, 1^2)$. Note that this p -value is valid on the condition that Z_E and Z_W are independent and identically distributed. However, a strict proof of their independence is difficult, because Z_E and Z_W are based on the regression analysis and rank transformation of Wilcoxon test. If (Z_E, Z_W) have the bivariate normal distribution and there is no correlation among these variables, Z_E and Z_W are independent. Therefore, we evaluated these two conditions via simulation studies.

Under the alternative hypothesis (i.e., there is a publication bias), both Z_E and Z_W are expected to be < 0 if the underlying treatment effect is positive ($\theta < 0$). However, if the underlying treatment effect is negative ($\theta > 0$), Z_W should be a value of > 0 , because the sample size of unpublished trials should be larger than that of published trials. In this case, the statistical power of the proposed method will decrease because Z_W has the opposite sign to Z_E . In order to address this problem, in our alternative approach, we use Z only if the median sample size of unpublished trials is smaller than that of published trials; otherwise, we use Z_E as a test statistic for the detection of publication bias.

5. Simulation Study

The aim of the simulation study was to evaluate and compare the performance of the established methods and the proposed method. For a range of events rates in the control group, treatment effects (log odds ratio), and number of trials in meta-analyses, we evaluated the type I error rates and the powers. Each simulation was repeated 10000 times. All the methods were conducted as one-sided tests with a 5 % level of significance.

5.1 Generation of trials

The underlying true treatment effects (θ) were log odds ratios of -0.6 , -0.4 , -0.2 , 0 , and 0.2 . The true treatment effect for the i th trial (θ_i) was generated according to a normal distribution with mean θ and heterogeneity variance τ^2 of 0 , 0.05 , and 0.25 (no, moderate, and large heterogeneity, respectively). The number of published trials (k) were varied as $k = 25$, 50 , and 75 trials according to the publication process described in the next section. To generate the total sample sizes of each trial (n_i), $\log n_i$ were generated according to a normal distribution with mean 3.798 and variance 1.104 , which were estimated from a sample of 1555 randomized controlled trials extracted from eight German journals (Schwarzer et al., 2002). The total sample size was rounded to the next even number and required at least 30 patients. The underlying outcome proportion in the control group (p_C) was generated according to a uniform distribution between 0.1 and 0.5 . Once the p_C had been determined, the corresponding underlying outcome proportion in the treatment group (p_T) was calculated from assumed log odds ratio θ_i . Then, we simulated the number of events in the treatment group (a_i) and control group (c_i) of each

trial from $\text{Bin}(n_i/2, p_T)$ and $\text{Bin}(n_i/2, p_C)$, respectively. We added 0.5 to all the cells in the 2×2 tables for all the trials in order to reduce small sample bias in the estimated log odds ratio and improve the asymptotic property of the standard error estimate.

5.2 Publication process

The publication status of each study (i.e., whether published or unpublished) was generated according to Bernoulli distribution with the probability of publication (π_i). For evaluating the type I error rate, π_i of 0.3, 0.5, and 0.7 were adopted under the assumption that π_i was constant with the result of each trial. To evaluate the power, we assumed that π_i depended on one-sided p -values of each trial: $p_i = \Phi(z_i)$, where $z_i = \hat{\theta}_i/s_i$. The probability of publication π_i was given by the following cumulative logistic function as the monotonic decrease function of z_i (Matsuoka et al., 2007),

$$\pi_i = \frac{1}{1 + \exp(-a - bz_i)}. \quad (9)$$

We set $a = -1.4$ and $b = -1$ so that the ratio of the probability of publication of the trial with a significant result to the probability of publication of the trial without a significant result was approximately 3. This rationale for setting the values of a and b came from the surveys to evaluate the relationship between publication status and statistical significance (Easterbrook et al., 1991; Dickersin et al., 1992; Sterling et al., 1995). The generation of a simulated trial was iterated until the number of published trials reached the given k . Therefore, the number of unpublished trials (l) was calculated by deducting k from the number of trials generated.

6. Results

6.1 Publication bias absent

Table 2 presents the proportion of meta-analyses identified as showing significant publication bias by each method for $\pi_i = 0.5$ in the absence of publication bias when there is no heterogeneity. A tendency of the inflation of type I error is apparent with increasing treatment effect, increasing number of trials in meta-analysis.

Begg's rank correlation method shows considerable inflation of type I error rate and exceeds the nominal level of significance (5%) in all situations (ranging from 9.6% to 33.9%). Macaskill's regression method is the most conservative and maintains the nominal level of significance in several situations (ranging from 3.6% to 6.8%). Egger's regression method is slightly more liberal than Macaskill's when the treatment effect is small. However, when the treatment effect is large, the degree of the inflation of type I error of Egger's regression method is larger than that of Macaskill's. The proposed method, in which Z is consistently used as the test statistic (denoted by "Proposed" in Table 2), is more conservative than Egger's regression method in all the situations except for $\theta = 0.2$. Meanwhile, the alternative proposed method, in which the test statistic is chosen as either Z or Z_E on the basis of the comparison of the median sample size between the published and unpublished trials (denoted by "Proposed (Choice)" in Table 2),

Jpn J Biomet Vol.32, No.1, 2011

shows further inflation type I error and about 3% more than Egger's regression method.

Table 3 shows that the type I error rate for $\pi_i = 0.5$ and $k = 50$ when there is moderate ($\tau^2 = 0.05$) or large ($\tau^2 = 0.25$) heterogeneity. As compared to the situation of no heterogeneity,

Table 2. Simulation results in the absence of publication bias; proportion of tests with significant result ($p < 0.05$, one-sided) when there is no heterogeneity; number of published trials ranged from 25 to 75; log odds ratio ranged from -0.6 to 0.2 ; probability of publication was 0.5 .

Heterogeneity	k	Method	log odds ratio (θ)				
			-0.6	-0.4	-0.2	0.0	0.2
$\tau^2 = 0$	25	Begg	0.211	0.181	0.160	0.124	0.096
		Egger	0.102	0.083	0.072	0.054	0.046
		Macaskill	0.057	0.057	0.050	0.048	0.046
		Proposed	0.083	0.073	0.064	0.054	0.045
		Proposed (Choice)	0.129	0.113	0.096	0.079	0.066
	50	Begg	0.295	0.246	0.200	0.154	0.124
		Egger	0.122	0.097	0.074	0.054	0.041
		Macaskill	0.064	0.056	0.052	0.049	0.040
		Proposed	0.094	0.079	0.067	0.052	0.044
		Proposed (Choice)	0.152	0.122	0.099	0.076	0.062
	75	Begg	0.339	0.290	0.240	0.182	0.129
		Egger	0.144	0.112	0.077	0.055	0.038
		Macaskill	0.068	0.060	0.052	0.044	0.036
		Proposed	0.107	0.088	0.067	0.055	0.042
		Proposed (Choice)	0.173	0.136	0.102	0.079	0.060

"Proposed" denotes the proposed method in which Z is consistently used as the test statistic. "Proposed (Choice)" denotes the proposed method in which the test statistic was chosen as either Z or Z_E on the basis of the comparison of the median sample size between the published and unpublished trials.

Table 3. Simulation results in the absence of publication bias; proportion of tests with significant result ($p < 0.05$, one-sided) when there is moderate ($\tau^2 = 0.05$) and large ($\tau^2 = 0.25$) heterogeneity; number of published trials was 50; log odds ratio ranged from -0.6 to 0.2 ; probability of publication was 0.5 .

Heterogeneity	k	Method	log odds ratio (θ)				
			-0.6	-0.4	-0.2	0.0	0.2
$\tau^2 = 0.05$	50	Begg	0.311	0.255	0.203	0.133	0.098
		Egger	0.173	0.136	0.111	0.085	0.068
		Macaskill	0.108	0.100	0.098	0.090	0.086
		Proposed	0.128	0.100	0.087	0.072	0.058
		Proposed (Choice)	0.199	0.160	0.136	0.107	0.087
$\tau^2 = 0.25$	50	Begg	0.439	0.349	0.268	0.189	0.122
		Egger	0.301	0.248	0.203	0.156	0.120
		Macaskill	0.153	0.142	0.148	0.134	0.138
		Proposed	0.200	0.167	0.144	0.115	0.090
		Proposed (Choice)	0.317	0.265	0.225	0.177	0.142

"Proposed" denotes the proposed method in which Z is consistently used as the test statistic. "Proposed (Choice)" denotes the proposed method in which the test statistic was chosen as either Z or Z_E on the basis of the comparison of the median sample size between the published and unpublished trials.

the inflation of type I error rate is large for all the methods. This is because an asymmetric funnel plot arises from not only publication bias but also heterogeneity, as described in the articles (Egger et al., 1997; Rothstein et al., 2005).

6.2 Publication bias present

Figure 5 presents the proportion of meta-analyses identified as showing significant publication bias by each method in the presence of publication bias when there is no heterogeneity. The power of all the methods increases with an increase in the number of trials in meta-analysis. This tendency is especially apparent in all the methods other than Begg's rank correlation method.

As regards Macaskill's regression method, the power is insufficient in all the situations. This

Fig. 5. Simulation results in the presence of publication bias; proportion of tests with a significant result ($p < 0.05$, one-sided) when there is no heterogeneity; number of published trials ranges from 25 to 75; log odds ratio ranges from -0.6 to 0.2 . “Proposed” denotes the proposed method in which Z is consistently used as the test statistic. “Proposed (Choice)” denotes the proposed method in which the test statistic was chosen as either Z or Z_E on the basis of the comparison of the median sample size between the published and unpublished trials.

is consistent with the results shown in the previous subsection. While Begg's rank correlation method is less affected by the size of the treatment effect, its power is generally low. Although the power of Egger's regression method is lower than that of Begg's rank correlation method for $k = 25$, it makes a comeback for $k = 50$ and 75 . The proposed method in which Z is consistently used as the test statistic (denoted by "Proposed" in Figure 5) is more powerful than Egger's regression method when the treatment effect is large. However, the power significantly decreases with a decrease in the size of the treatment effect. On the other hand, the alternative proposed method in which the test statistic is chosen as either Z or Z_E (denote by "Proposed (Choice)" in Figure 5) is more powerful or comparable to Egger's regression method in all the situations. The reason the "Proposed" method shows insufficient power when the treatment effect is small and negative is that the median sample size of unpublished trials must be large as compared to that of published trials.

Figure 6 presents the power of each method for $k = 50$ when there is moderate ($\tau^2 = 0.05$) and large ($\tau^2 = 0.25$) heterogeneity. As compared to the situations without heterogeneity, the power is slightly decreased and the influence of the treatment effect becomes large in all the methods.

As described in the previous section, the inflation of type I error in each method was shown in many situations. As the result of the trade-off relationship between type I error and type II error, statistical tests exceeding the nominal significant level provides the larger power than that

Fig. 6. Simulation results in the presence of publication bias; proportion of tests with significant result ($p < 0.05$, one-sided) when there is moderate ($\tau^2 = 0.05$) and large ($\tau^2 = 0.25$) heterogeneity; number of published trials ranges from 25 to 75; log odds ratio ranges from -0.6 to 0.2 . "Proposed" denotes the proposed method in which Z is consistently used as the test statistic. "Proposed" denotes the proposed method in which Z is consistently used as the test statistic. "Proposed (Choice)" denotes the proposed method in which the test statistic was chosen as either Z or Z_E on the basis of the comparison of the median sample size between the published and unpublished trials.

of the tests keeping nominal significant level. Therefore, the power of each method was compared using the adjusted critical value to maintain each type I error at approximately 5%. Although the power of each method decreased by a little fewer than 10%, the relative power relationship among methods was not changed.

6.3 Comparison of the sample size between published and unpublished trials

Figure 7 presents the median sample size of published and unpublished trials in simulated meta-analyses for (a) $\pi_i = 0.5$, $k = 50$, and moderate heterogeneity in the absence of publication bias and (b) $k = 50$, and moderate heterogeneity in the presence of publication bias. Similar results are shown in other situations. In the absence of publication bias, the sample sizes are comparable in published and unpublished trials because the probability of publication of each trial is constant with p -value, that is, it does not depend on the sample size.

Fig. 7. One of the simulation results; median sample size of published and unpublished trials. (a) $pp_i = 0.5$, $k = 50$, moderate heterogeneity in the absence of publication bias; (b) $k = 50$, moderate heterogeneity in the presence of publication bias.

In the presence of publication bias, the sample size of published trials decreases and the sample size of unpublished trials increases with decreasing treatment effect. When the treatment effect is positive ($\theta < 0$), the sample size of published trials is larger than that of unpublished trials. However, this tendency is reversed when the treatment effect is negative ($\theta > 0$). This is the reason that the power of the proposed method in which Z is consistently used decreases with a decrease in the treatment effect.

6.4 Characteristics of Z_E and Z_W

As mentioned in Section 4., the p -value based on Z in the proposed method is valid on the condition that Z_E and Z_W are independently and identically distributed. We evaluated whether (Z_E, Z_W) have the bivariate normal distribution and there is no correlation between Z_E and Z_W via simulation studies.

Table 4 shows joint probabilities that Z_E and Z_W fell within one standard deviation and over in the absence of publication bias in order to compare the theoretical values of the bivariate

Table 4. Joint Probability of Z_E and Z_W in the simulation; $\theta = -0.2$, $\pi_i = 0.5$, $k = 50$, no heterogeneity in the absence of publication bias.

$Z_E \backslash Z_W$	$(-\infty, -1]$	$(-1, 0]$	$(0, 1]$	$(1, \infty]$
$(1, \infty]$	0.033 (0.025)	0.062 (0.054)	0.046 (0.054)	0.019 (0.025)
$(0, 1]$	0.070 (0.054)	0.126 (0.117)	0.010 (0.117)	0.043 (0.054)
$(-1, 0]$	0.069 (0.054)	0.121 (0.117)	0.104 (0.117)	0.045 (0.054)
$(-\infty, -1]$	0.035 (0.025)	0.056 (0.054)	0.053 (0.054)	0.019 (0.025)

top: proportion of simulation results which were fell within each range

(bottom): theoretical values

standard normal distribution with independent marginals. Basically, simulation results are in agreement with the theoretical values and suggest that (Z_E, Z_W) have the bivariate standard normal distribution with independent marginals. Moreover, figure 8 shows the normal Q-Q plot of the test statistics $Z = (Z_E + Z_W)/\sqrt{2}$ in the simulation. The linearity of the points suggests that the data are normally distributed.

Fig. 8. Normal Q-Q plot of the test statistics $Z = (Z_E + Z_W)/\sqrt{2}$ in the simulation; $\theta = -0.2$, $\pi_i = 0.5$, $k = 50$, no heterogeneity in the absence of publication bias.

In addition, we evaluated the correlation between Z_E and Z_W in the simulation studies. The correlations between Z_E and Z_W are vanishingly small, ranging from -0.02 to 0.02 in the absence of publication bias and from -0.03 to 0.02 in the presence of publication bias. Figure 9 presents the scatter plots of Z_E and Z_W . No correlation between Z_E and Z_W can be visually confirmed from these scatter plots.

A strict proof of independence is quite difficult because Z_E is based on rank-transformed data. The essential idea of the proposed method is to improve the power to detect publication

Fig. 9. One of the simulation results; scatter plots of Z_E and Z_W . (a) $\theta = -0.2$, $pp_i = 0.5$, $k = 50$, no heterogeneity in the absence of publication bias, (b) $\theta = -0.2$, $k = 50$, no heterogeneity in the presence of publication bias.

bias with the information on sample size of unpublished trials using the clinical trial registration system. The three commonly used methods are based on the correlation (i.e., regression slope) between the estimated treatment effect and its precision in published trials. In contrast, the registry system makes it possible to compare the precision (i.e., sample size) between published and unpublished trials. Assuming independence, this article proposed statistics combining the correlation and location shift in two-sample problems. As described above, this assumption seems to be valid. While strict proof of independence is quite difficult, it should be noted that the independence between the slope of the regression of the estimated treatment effect on its precision in published trials and the difference of the mean precision between published and unpublished trials can be proved as follows:

Let y_i denote the estimated treatment effect and x_i denote its precision. The slope (β) of regression of y_i on x_i is given as

$$\begin{aligned}\hat{\beta} &= \frac{S_{xy}}{S_{xx}} \\ &= S_{xx}^{-1} \left\{ \sum_{i=1}^k (y_i - \bar{y}) x_i \right\} \\ &= S_{xx}^{-1} \left\{ \sum_{i=1}^{k+l} (y_i - \bar{y}) x_i \right\} \\ &= S_{xx}^{-1} \mathbf{E}^T \mathbf{X},\end{aligned}\tag{10}$$

where $\mathbf{E} = (y_1 - \bar{y}, \dots, y_k - \bar{y}, 0, \dots, 0)^T$, $\mathbf{X} = (x_1, \dots, x_k, x_{k+1}, \dots, x_{k+l})$, and $S_{xx}^{-1} = \sum_{i=1}^k (x_i - \bar{x})^2$. At the same time, the difference of the mean precision between published and unpublished trials is given by $\mathbf{W}^T \mathbf{X}$, where $\mathbf{W} = (-1/k, \dots, -1/k, -1/l, \dots, -1/l)^T$. These coefficient vectors for $\mathbf{X}$ ($\mathbf{E}$ and $\mathbf{W}$) are mutually orthogonal because the inner product is 0. Thus, the slope

and the difference of the mean precision are independent of each other. In addition to the simulation results, the validity of the assumption of independence between the correlation and location shift statistics is suggested.

We also evaluated the contributions of Z_E and Z_W to Z . In the absence of publication bias, Z_W and Z_E equally contributed to Z . In the presence of publication bias, the contributions of Z_W and Z_E were comparable for $\theta = -0.6$, and the contribution of Z_W decreased and the contribution of Z_E increased with decreasing treatment effect (For $\theta = 0.2$, the contribution of Z_E and Z_W were 15% and 85%, respectively). This was because the difference of sample sizes between published and unpublished trials became smaller with decreasing treatment effect as shown in Figure 7.

7. A Case Study

It is usual that almost 5 years is required between the planning of a clinical trial and its completion. Moreover, with considering the duration until publication, the number of trials in publication phase may not be sufficient at this time. Therefore, it is difficult to show an actual case study for the proposed method using the clinical trial registration system in this paper. Instead, a hypothetical case study, which is based on actual meta-analysis targeting both published and unpublished trials, is presented in this section.

Nissen and Wolski (2007) conducted the meta-analysis to assess the effect of rosiglitazone, which was widely used to treat patients with type 2 diabetes mellitus, on the risk of myocardial infarction and death from cardiovascular causes. A total of 42 trials were included in the meta-analysis. Of these trials, the 15 trials were published in medical journals and the remaining 27 trials were not published. However, they obtained the results of unpublished trials from the GlaxoSmithKline Web site. Since the overall trial population including unpublished trials were identified in the meta-analysis, this can be considered as a similar context in the proposed method. As the result of assessing publication bias using the three established methods and the proposed method (significance level: 5%), only the proposed method could detect the publication bias, while none of the established methods could detect the publication bias (Table 5).

Table 5. Result of Case Study

Method	p -value (one-sided)
Begg	0.299
Egger	0.100
Macaskill	0.546
Proposed	0.042
Proposed (Choice)	0.042

8. Discussion

The goal of this paper is to improve the statistical power for detecting publication bias, compared to the commonly used methods. To achieve this goal, information on the target sample sizes of all registered trials including unpublished trials was utilized in the proposed method.

From the simulation results, the proposed method in which the test statistic was chosen as either Z or Z_E ("Proposed (Choice)" method) was generally the most powerful method, while the type I error was slightly inflated compared to other methods excluding Begg's rank correlation method. Moreover, only the proposed method can detect publication bias in the case study. Therefore, the proposed method appeared to perform well as compared with the other method.

However, there are some remaining points to be considered. We assumed that actual sample size is similar to target sample size in this paper. In fact, however, actual sample size is sometimes differently smaller than target sample size because of the difficulty of patient recruitment and so on. If the percentage of target sample size is consistent among clinical trials (e.g., actual sample size is 80% of target sample size in all clinical trials), the performance of the proposed method is not affected because the degrees of lack of sample size is balanced between published and unpublished trials. On the other hand, if the percentage of actual sample size in target sample size varies among clinical trials, the performance of the proposed method could be affected as follows:

1. If publication bias is present (under alternative hypothesis) and true treatment effect is positive, clinical trials in which target sample size can not be achieved less likely to be published due to the lack of statistical power. Therefore, the difference of target sample size between published and unpublished trials is smaller than that of actual sample size. As a result, the power of the proposed method decreases.
2. If publication bias is present (under alternative hypothesis) and true treatment effect is negative, clinical trials in which target sample size can not be achieved more likely to be published because a significant result may be demonstrated by chance. Therefore, the difference of target sample size between published and unpublished trials is larger than that of actual sample size. As a result, the power of the proposed method increases.
3. If publication bias is absent (under null hypothesis), the degrees of unreached sample size is balanced between published and unpublished trials because results of clinical trials are published independently of a result itself. As a result, the type I error of the proposed method is not affected.

Anyway, when the proposed method is applied, we should confirm similarity between target and actual sample size in published trials. It is also important to evaluate how much the difference between actual and target sample size affect the power of the proposed method in the future research issue.

In order to implement the proposed method, some additional works are needed owing to

using information from clinical trial registration system. When conducting a meta-analysis as usual, clinical trials which meet the pre-specified criteria (e.g., patient population, intervention (and control), endpoints) for inclusion in a meta-analysis are selected from published information (e.g., medical journals). In addition to above, all clinical trials which meet the same criteria have to be extracted from the clinical trial registration system, and these registered trials have to be cross-checked with the published trials to identify the publication status. Considering that the clinical trial registration will be more widespread in the near future, all clinical trials will be searched from the registration in order to grasp whole population of clinical trials and prevent to miss some published trials even when just conducting a meta-analysis. Therefore, such additional workload due to implementing the proposed method will not be heavy.

REFERENCES

- Begg, C.B. and Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics* **50**, 1088–1101.
- DeAngelis, C.D., Drazen, J.M., Frizelle, F.A., Haug, C., Hoey, J., Horton, R., Kotzin, S., Laine, C., Marusic, A., Overbeke, A.J.P.M., Schroeder, T.V., Sox, H.C., and Van Der Weyden, M.B. (2004). Clinical trial registration: a statement from the international committee of medical journal editors. *Journal of the American Medical Association* **292**, 1363–1364.
- Dickersin, K., Min, Y.I., and Meinert, C.L. (1992). Factors influencing publication of research results: follow-up of applications submitted to two institutional review boards. *Journal of the American Medical Association* **267**, 374–378.
- Easterbrook, P.J., Berlin, J.A., Gopalan, R., and Matthews, D.R. (1991). Publication bias in clinical research. *Lancet* **337**, 867–872.
- Egger, M., Smith, G.D., Schneider, M., and Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal* **315**, 629–634.
- Galbraith, R.F. (1998). A note on graphical presentation of estimated odds ratios from several clinical trials. *Statistics in Medicine* **7**, 889–894.
- Macaskill, P., Walter, S.D., and Irwig, L. (2001). A comparison of methods to detect publication bias in meta-analysis. *Statistics in Medicine* **20**, 641–654.
- Matsuoka, N., Hasegawa, C., and Hamada, C. (2007). A practical method adjusting for publication bias in meta-analysis based on p-value. *Japanese Journal of Biometrics* **28**, 19–36.
- Nissen, S.E. and Wolski, K.W. (2007). Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *The New England Journal of Medicine* **356**, 2457–2471.
- Rothstein, H.R., Sutton, A.J., and Borenstein, M. (2005). Publication bias in meta-analysis: prevention, assessment and adjustments. *Wiley*.

- Schwarzer, G., Antes, G., and Schumacher, M. (2002). Inflation of type I error rate in two statistical tests for the detection of publication bias in meta-analyses with binary outcomes. *Statistics in Medicine* **21**, 2465–2477.
- Song, F., Eastwood, A.J., Gilbody, S., Duley, L., and Sutton, A.J. (2000). Publication and related bias. *Health Technology Assessment* **4**.
- Sterling, T.D., Rosenbaum, W.L., and Weinkam, J.J. (1995). Publication decisions revisited: the effect of the outcome of statistical tests on the decision to publish and vice versa. *The American Statistician* **49**, 108–112.
- Terrin, N., Schmid, C.H., and Lau, J. (2005). In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias. *Journal of Clinical Epidemiology* **58**, 894–901.
- Thornton, A. and Lee, P. (2000). Publication bias in meta-analysis: its cause and consequences. *Journal of Clinical Epidemiology* **53**, 207–216.
- World Health Organization. WHO Trial Registration Data Set.
<http://www.who.int/ictrp/network/trds/en/index.html>. Accessed 23 May 2009.