Adjusting for publication bias in meta-analysis via inverse probability weighting using clinical trial registries

Ao Huang | Kosuke Morikawa | Tim Friede | Satoshi Hattori

Abstract
Publication bias is a major concern in conducting systematic reviews and meta-analyses. Various sensitivity analysis or bias-correction methods have been developed based on selection models, and they have some advantages over the widely used trim-and-fill bias-correction method. However, likelihood methods based on selection models may have difficulty in obtaining precise estimates and reasonable confidence intervals, or require a rather complicated sensitivity analysis process. Herein, we develop a simple publication bias adjustment method by utilizing the information on conducted but still unpublished trials from clinical trial registries. We introduce an estimating equation for parameter estimation in the selection function by regarding the publication bias issue as a missing data problem under the missing not at random assumption. With the estimated selection function, we introduce the inverse probability weighting (IPW) method to estimate the overall mean across studies. Furthermore, the IPW versions of heterogeneity measures such as the between-study variance and the $I^2$ measure are proposed. We propose methods to construct confidence intervals based on asymptotic normal approximation as well as on parametric bootstrap. Through numerical experiments, we observed that the estimators successfully eliminated bias, and the confidence intervals had empirical coverage probabilities close to the nominal level. On the other hand, the confidence interval based on asymptotic normal approximation is much wider in some scenarios than the bootstrap confidence interval. Therefore, the latter is recommended for practical use.

Keywords
clinical trial registry, missing not at random, propensity score, sensitivity analysis, systematic review

1 | INTRODUCTION

Meta-analyses play a very important role in medical research and may have substantial impacts in establishing sound medical evidence. Meta-analysts try to gather all the available evidence by conducting systematic literature searches including not only the scientific literature but also the so-called grey literature such as documents for regulation of new drug applications and conference abstracts (Gopalakrishnan & Ganeshkumar, 2013). Despite
such painstaking efforts, it is very difficult to collect all information; thus, reporting bias may arise when some negative results are not reported by investigators or are not likely to be accepted by scientific journals or might be presented in a way that they become positive. Especially when it comes to the situation that publication status (publication or non-publication) depends on the nature and the direction of research findings, it is usually referred to as the publication bias (Thornton & Lee, 2000).

The funnel plot and the trim-and-fill method are among the most widely used methods to identify and adjust for publication bias (Duval & Tweedie, 2000; Egger et al., 1997). Despite their simple interpretability through graphical presentation, results obtained via these methods may be misleading (Peters et al., 2007; Terrin et al., 2003). Modeling the selective publication process by a selection model may yield more reliable and interpretable results to quantify the impact of publication bias (Carpenter et al., 2009; Schwarzer et al., 2010). The Copas–Shi selection model was suggested to be preferable to the trim-and-fill method by Schwarzer et al. (2010). It is an adaptation of the Heckman selection model, which was first proposed in the context of econometrics, and was then introduced to the area of meta-analysis by Copas (1999) and Copas and Shi (2000). A notable feature of the Copas–Shi selection model is that it models the selection process based on a simple Gaussian latent variable, which enables us to address the publication bias issue in the sensitivity analysis of various kinds of meta-analysis, such as the meta-analysis of diagnostic studies (Hattori & Zhou, 2018; Piao et al., 2019) and network meta-analysis (Mavridis et al., 2013; Marks-Anglin et al., 2022). However, the interpretation of Heckman-type selection functions might be unsatisfactory in medical research. Selection functions defined with the test statistics used in each publication might be more appealing, since P-values might be a highly influential factor for the decision to publish. Preston et al. (2004) discussed the maximum conditional likelihood estimation with a series of one-parameter selection functions based on the P-values; Copas (2013) proposed a likelihood-based sensitivity analysis method with the selection function modeling the t-type statistics directly. In this paper, we denote these selection functions as t-type selection functions.

Registration of study protocols to clinical trial registries is a non-statistical approach to prevent selective publication; by prospectively registering all the clinical trials, one can identify all the studies and then address whether selective publication matters or not. According to the recommendation by the International Committee of Medical Journal Editors (ICMJE) (DeAngelis et al., 2005), several clinical trial registries, such as ClinicalTrials.gov, World Health Organization’s (WHO) International Clinical Trials Registry Platform (ICTRP), EU Clinical Trials Register (EUCTR) and International Standard Randomized Controlled Trial Number (ISRCTN), have been established and are widely used in practice (for the links of these clinical trial registries, see Web Table 36 in Web Appendix G). Potentially, the accumulated information in clinical trial registries could be very useful in reducing publication bias (Baudard et al., 2017; Hart et al., 2012). However, their roles in the current meta-analysis practice are usually limited as a searching tool to identify conducted but still unpublished studies. Information (e.g., the planned sample sizes) in the clinical trial registries has not been efficiently utilized, in particular, to address the potential impacts of selective publication on the estimation of the effect size.

In this paper, we develop a simple inference procedure to correct for publication bias under the selective publication process driven by the statistical significance of the result, more specifically, the t-type statistic of each study, which is an appealing alternative to the Heckman-type selection function by Copas and Shi (2000). We propose a publication bias adjusted estimator based on the inverse probability weighting (IPW), which is a widely used technique in missing data problems and causal inference. Considering the correspondence between the propensity score in missing data and causal inference and the selection function in meta-analysis, application of the IPW idea in meta-analysis is very natural and indeed is not new. Matsuoka et al. (2007) and Mathur and VanderWeele (2020) examined the performance of the IPW estimator on quantifying publication bias in the meta-analysis context. However, both methods relied on sensitivity analysis approaches. That is, the publishing probability, which corresponds to the propensity score in the IPW estimator, was fixed and was not estimated from data. This can be a very difficult task in practice. With the planned sample size in the clinical trial registries, we introduce an estimating equation for unknown parameters in the selection function, borrowing the idea to handle the propensity score in the general missing data problem under the missing not at random assumption (Kott & Chang, 2010; Miao & Tchetgen Tchetgen, 2016; Morikawa & Kim, 2021). The estimating equation is tractable and once the parameters in the selection function are estimated, our IPW estimator for the overall mean over studies is a very simple closed form expression. In addition to providing a combined mean, evaluation of the between-study heterogeneity is also an important objective of meta-analyses; the common-effect assumption is implausible in many systematic reviews and therefore random-effect models are recommended in practice (Borenstein et al., 2010). We propose an IPW-type DerSimonian–Laird estimator for the between-study variance and also some other heterogeneity measures, all of which have a simple closed form. We show consistency and asymptotic normality of the proposed estimators and
propose a method to construct confidence intervals (CIs) for the overall mean and the between-study variance based on a consistent estimator of the asymptotic variance. An alternative construction of the CIs is also proposed with parametric bootstrapping.

2 | BASIC SETUP AND THE STANDARD METHODS FOR META-ANALYSIS

Suppose we are conducting a meta-analysis of \( N \) published studies to compare two treatment groups. Let the estimated treatment effect of the \( i \)-th study denoted by \( y_i \) such as the log-odds ratio or the log-hazard ratio, and its standard error \( \sigma_i \) is supposed to be available. Following the standard convention in the meta-analysis field, \( \sigma_i \) is assumed to be known in theoretical development. We suppose the following random-effects model; given \( \mu_i \) and \( \sigma_i, y_i \sim N(\mu_i, \sigma_i^2) \). Here, \( \mu_i \) is the true value of the \( i \)-th study and is regarded as a random-effect such that \( \mu_i \sim N(\mu, \tau^2) \), where \( \mu \) is the overall treatment effect and \( \tau^2 \) is the unknown between-study variance. Then, the marginal model \( y_i \sim N(\mu, \sigma_i^2 + \tau^2) \) follows from the above.

The inverse variance weighted estimator (Cochran, 1954) for \( \mu \) is denoted by

\[
\hat{\mu} = \frac{\sum_{i=1}^{N} \omega_i y_i}{\sum_{i=1}^{N} \omega_i},
\]

where \( \omega_i = (\sigma_i^2 + \tau^2)^{-1} \). In practice, \( \tau^2 \) should be estimated and various estimators are available. In this paper, we consider the DerSimonian–Laird (DL) estimator (DerSimonian & Laird, 1986), which is given by

\[
\tau^2_{DL} = \max \left\{ 0, \frac{Q - (N - 1)}{\sum_{i=1}^{N} \sigma_i^{-2} - \sum_{i=1}^{N} \sigma_i^{-4} / \sum_{i=1}^{N} \sigma_i^{-2}} \right\},
\]

where \( Q = \sum_{i=1}^{N} (y_i - \hat{\mu}_i)^2 / \sigma_i^2 \) is Cochran’s \( Q \) statistic. \( \hat{\mu}_i \) is the fixed-effect estimator, which is defined by Equation (1) with \( \tau^2 = 0 \).

3 | PROPOSED METHOD

3.1 | Clinical trial registry

In addition to \( N \) published studies, suppose we identify \( M \) unpublished studies by using clinical trial registries. For \( i = 1, 2, \ldots, N + M \), let the random variable \( D_i \) be 1 if the \( i \)-th study is published and be 0 if unpublished. Without loss of generality, we assume that the first \( N \) studies are published. As defined in Section 2, for published studies, \((y_i, \sigma_i)\) are available. As argued in the introduction, for studies registered in a clinical trial registry, the planned sample sizes of the two groups (not separately by groups) are available regardless of clinical trial registry systems. Let \( n_i \) be the number of patients enrolled in the two groups for published studies and be the planned sample size in the two groups for unpublished studies. We assume that \( n_i \) is consistent with the actual sample size for unpublished studies. Then, we suppose the following data are available; for \( i = 1, 2, \ldots, N \) (published studies), \((y_i, \sigma_i, n_i)\) is available and for \( i = N + 1, \ldots, N + M \) (unpublished studies), only \( n_i \) is available. In the following, we suppose \((y_i, \sigma_i, n_i)\) for \( i = 1, 2, \ldots, N + M \) are random samples from a population.

3.2 | Selection functions based on \( t \)-type statistic

In this section, we introduce some selection functions describing selective publication processes. We focus on the selection functions defined with the \( t \)-type statistic \( t_i = y_i / \sigma_i \). Let \( \pi_i(\beta) = P(D_i = 1 \mid y_i, \sigma_i, n_i; \beta) \) denote the probability of the \( i \)-th study with \((y_i, \sigma_i, n_i)\) being published, where \( \beta \) is a parameter (vector). We consider one- or two-parameter selection functions. For two-parameter cases, we denote \( \beta = (\beta_0, \beta_1) \).

Preston et al. (2004) considered several one-parameter selection functions including the one-parameter logistic function

\[
\pi_i(\beta) = \frac{\exp(-\beta(1 - \Phi(t_i)))}{1 + \exp(-\beta(1 - \Phi(t_i)))},
\]

and the modified one-parameter logistic function

\[
\pi_i(\beta) = \frac{2 \exp(-\beta \sigma_i(1 - \Phi(t_i)))}{1 + \exp(-\beta \sigma_i(1 - \Phi(t_i)))},
\]

where \( \Phi(\cdot) \) is the cumulative function of the standard normal distribution. Other one-parameter selection models, such as the half-normal and the negative-exponential selection functions and their modified versions, were also considered. Preston et al. (2004) proposed to estimate all the parameters of \((\mu, \tau^2, \beta)\) by maximizing the conditional log-likelihood function for published studies (see details in Web Appendix A1). However, as they commented, parameters in the selection function might be estimated imprecisely, which in turn may influence the estimates of effect size and result in an unreasonable CI. Probably, due to difficulty in estimation, Preston et al. (2004) mainly focused on one-parameter selection functions. Although these one-parameter selection functions
have an advantage of simplicity, they have a disadvantage of impossibility to describe the publication process that does not depend on the $t$-type statistic, or say a random selection. If some studies are unpublished independently from outcomes, $\beta$ in the selection function (3) or (4) should be zero. Then, the marginal selection probability $p = P(D_i = 1)$ should be 1, which does not allow the existence of randomly unpublished studies.

Besides, two-parameter selection functions are also considered including the two-parameter probit model

$$\pi_i(\beta) = \Phi(\beta_0 + \beta_1 t_i),$$

and the two-parameter logistic model

$$\pi_i(\beta) = \frac{\exp(\beta_0 + \beta_1 t_i)}{1 + \exp(\beta_0 + \beta_1 t_i)}.$$  

Copas (2013) proposed a likelihood-based sensitivity analysis method; with the marginal selection probability $p$ fixed, one could estimate all the parameters by maximizing the observed conditional likelihood iteratively. Then, the impact of the publication bias can be studied by monitoring how the effect size changed as the selection probability $p$ decreased (see details in Web Appendix A2).

### 3.3 Inverse probability weighting method for publication bias adjustment

With publication indicator $D_i$, the estimator (1) is expressed as

$$\hat{\beta} = \frac{\sum_{i=1}^{N} \omega_i y_i^{*}}{\sum_{i=1}^{N} \omega_i} = \frac{\sum_{i=1}^{S} \omega_i D_i y_i^{*}}{\sum_{i=1}^{S} \omega_i D_i},$$

where $S = N + M$. This representation motivates us to use an estimator of the form

$$\hat{\beta}_{IPW}(\beta, \tau^2) = \frac{\sum_{i=1}^{S} \frac{1}{\sigma^2_i + \tau^2} \frac{D_i}{\pi_i(\beta)} y_i^{*}}{\sum_{i=1}^{S} \frac{1}{\sigma^2_i + \tau^2} \pi_i(\beta)}.$$  

This is a natural analogy of the inverse probability weighted (IPW) estimator by the propensity score, which is widely used in missing data problems and causal inference. For estimation of $\beta$, we consider the following estimating equation

$$U(\beta) = \sum_{i=1}^{S} \left( 1 - \frac{D_i}{\pi_i(\beta)} \right) g(n_i) = 0,$$  

where $g(n_i)$ is a function of the same dimension as $\beta$. This estimating equation is motivated by the propensity score analysis in the missing not at random setting (Kott & Chang, 2010; Miao & Tchetgen Tchetgen, 2016; Morikawa & Kim, 2021). Indeed, we could prove that,

$$E[U(\beta^*) | y_i, \sigma, n_i] = E \left[ \sum_{i=1}^{S} \left( 1 - \frac{D_i}{\pi_i(\beta^*)} \right) g(n_i) | y_i, \sigma, n_i \right] = \sum_{i=1}^{S} \left( 1 - \frac{\pi_i(\beta^*)}{\pi_i(\beta^*)} \right) g(n_i) = 0$$

holds if $\pi_i(\beta^*)$ is correctly specified, where $\beta^*$ is the true value of $\beta$. Regarding the specification of $g(n_i)$, one may make an efficiency argument (Morikawa & Kim, 2021), but we employ rather simple ones as follows. When we consider a one-parameter selection function such as Equations (3) and (4), we use

$$U(\beta) = \sum_{i=1}^{S} \left( 1 - \frac{D_i}{\pi_i(\beta)} \right) \sqrt{n_i} = 0.$$  

When we use a two-parameter selection function such as Equations (5) and (6), we consider the estimating equations,

$$U(\beta) = \sum_{i=1}^{S} \left( 1 - \frac{D_i}{\pi_i(\beta)} \right) \left( \frac{1}{\sqrt{n_i}} \right) = 0.$$  

The solution to Equation (11) or (12) is denoted by $\hat{\beta}$, and $\hat{\beta}$ is a consistent estimator to the true value of $\beta^*$ if the selection function is correctly specified (Kott & Chang, 2010; Miao & Tchetgen Tchetgen, 2016; Morikawa & Kim, 2021) (see the proof in Web Appendix B).

For the one-parameter selection functions, one can easily see that Equation (11) is a monotone function of $\beta$ and then the equation can be easily solved by the Newton–Raphson or the binary search methods. For the two-parameter selection functions, the Hessian matrix for Equation (12) may not be positive definite, and we observed computational difficulties when applying the Newton–Raphson method. We propose to obtain the solution to Equation (12) by minimizing

$$\sum_{i=1}^{S} \left| 1 - \frac{D_i}{\pi_i(\beta)} \right| + \sum_{i=1}^{S} \left| \frac{D_i}{\pi_i(\beta)} \right| \sqrt{n_i}.$$  

We use the nlminb() function in R (package stats, version 3.6.2) for implementation.
To estimate $\tau^2$, we propose an IPW version of the DL estimator, which is defined by $\hat{\tau}^2_{\text{IPW}} = \hat{\tau}^2_{\text{IPW}}(\hat{\beta})$, where

$$\hat{\tau}^2_{\text{IPW}}(\hat{\beta}) = \max \left\{ 0, \frac{Q_{\text{IPW}}(\hat{\beta}) - \{S - 1\}}{\sum_{i=1}^{S} \frac{1}{\sigma_i^2} \frac{1}{\pi_i(\hat{\beta})} - A_S(\hat{\beta})/B_S(\hat{\beta})} \right\},$$

(14)

$$A_S(\hat{\beta}) = S^{-1} \sum_{i=1}^{S} \frac{1}{\sigma_i^2} \frac{D_i}{\pi_i(\hat{\beta})}, B_S(\hat{\beta}) = S^{-1} \sum_{i=1}^{S} \frac{1}{\sigma_i^2} \frac{D_i}{\pi_i(\hat{\beta})},$$

$$Q_{\text{IPW}}(\hat{\beta}) = \sum_{i=1}^{S} \frac{1}{\sigma_i^2} \frac{D_i}{\pi_i(\hat{\beta})} \left( y_i - \mu_{\text{IPW}}(\beta) \right)^2,$$

(15)

and

$$\mu_{\text{IPW}}(\beta) = \frac{\sum_{i=1}^{S} \frac{1}{\sigma_i^2} \frac{D_i}{\pi_i(\hat{\beta})} y_i}{\sum_{i=1}^{S} \frac{1}{\sigma_i^2} \frac{D_i}{\pi_i(\hat{\beta})}}.$$  

(16)

We call the estimator (14) the IPW-DL estimator. $Q_{\text{IPW}}(\beta)$ and $\mu_{\text{IPW}}(\beta)$ are the IPW versions of $Q$ statistics in Equation (2) and the fixed-effect model estimator, respectively.

Finally, we propose the IPW estimator $\hat{\mu}_{\text{IPW}} = \hat{\mu}_{\text{IPW}}(\beta, \hat{\tau}^2_{\text{IPW}})$ for $\mu$. In Web Appendix B, we show the consistency of $\hat{\mu}_{\text{IPW}}$ and $\hat{\tau}^2_{\text{IPW}}$ if the selection function is correctly specified as $S$ goes to infinity and $n_i$ goes to infinity for each $i$. The CIs for $\mu$, $\tau^2$ as well as $\hat{\beta}$, can be constructed based on the asymptotic normality of $(\hat{\mu}, \hat{\tau}^2, \hat{\beta})$ with a consistent estimator of their asymptotic variance, whose derivations and definitions are given in Web Appendix C.

### 3.4 Parametric bootstrap confidence intervals

Alternatively, a parametric bootstrap approach may be used to construct CIs. Conditional on the data, parametric bootstrap samples $\tilde{y}_i$ are generated from $\tilde{y}_i \sim N(\mu_{\text{IPW}}, \sigma_i^2 + \hat{\tau}^2_{\text{IPW}})$ (Turner et al., 2000; Viechtbauer, 2007). Define

$$\tilde{\mu}(\beta) = \sum_{i=1}^{S} \left\{ 1 - \frac{D_i}{\pi_i(\hat{\beta})} \right\} \frac{1}{\sigma_i^2} \frac{1}{\pi_i(\hat{\beta})} g(n_i) = 0,$$

(17)

where $\pi_i(\hat{\beta})$ is defined by $\pi_i(\beta)$ replacing $t_i = y_i/\sigma_i$ with $\tilde{y}_i/\sigma_i$. Let the solution to $\tilde{\mu}(\beta) = 0$ denoted by $\tilde{\beta}$. Define $\tilde{\tau}^2_{\text{IPW}} = \tilde{\tau}^2_{\text{IPW}}(\tilde{\beta})$, where

$$\tilde{\tau}^2_{\text{IPW}}(\tilde{\beta}) = \max \left\{ 0, \frac{\tilde{Q}_{\text{IPW}}(\tilde{\beta}) - \{S - 1\}}{\sum_{i=1}^{S} \frac{1}{\sigma_i^2} \frac{1}{\pi_i(\tilde{\beta})} - A_S(\tilde{\beta})/B_S(\tilde{\beta})} \right\}.$$  

(18)

$$A_S(\tilde{\beta}) = S^{-1} \sum_{i=1}^{S} \frac{1}{\sigma_i^2} \frac{D_i}{\pi_i(\tilde{\beta})}, B_S(\tilde{\beta}) = S^{-1} \sum_{i=1}^{S} \frac{1}{\sigma_i^2} \frac{D_i}{\pi_i(\tilde{\beta})},$$

$$\tilde{Q}_{\text{IPW}}(\tilde{\beta}) = \sum_{i=1}^{S} \frac{1}{\sigma_i^2} \frac{D_i}{\pi_i(\tilde{\beta})} \left( \tilde{y}_i - \mu_{\text{IPW}}(\beta) \right)^2,$$  

(19)

and

$$\tilde{\mu}_{\text{IPW}}(\tilde{\beta}) = \frac{\sum_{i=1}^{S} \frac{1}{\sigma_i^2} \frac{D_i}{\pi_i(\tilde{\beta})} \tilde{y}_i}{\sum_{i=1}^{S} \frac{1}{\sigma_i^2} \frac{D_i}{\pi_i(\tilde{\beta})}}.$$  

(20)

Then, define $\tilde{\mu}_{\text{IPW}} = \tilde{\mu}_{\text{IPW}}(\tilde{\beta}, \tilde{\tau}^2_{\text{IPW}})$, where

$$\tilde{\mu}_{\text{IPW}}(\tilde{\beta}, \tilde{\tau}^2_{\text{IPW}}) = \frac{\sum_{i=1}^{S} \frac{1}{\sigma_i^2} \frac{D_i}{\pi_i(\tilde{\beta})} \tilde{y}_i}{\sum_{i=1}^{S} \frac{1}{\sigma_i^2} \frac{D_i}{\pi_i(\tilde{\beta})}}.$$  

(21)

For $i = 1, 2, \ldots, N$, a sufficiently large number (say, 1000) of parametric bootstrap samples of $\tilde{y}_i$ are generated. Let the number of bootstrap samples be denoted by $B$ and the $b$th bootstrap sample be denoted by $\tilde{y}_i^{(b)}$. We denote $\tilde{\mu}_{\text{IPW}}$ with the $b$th bootstrap sample by $\tilde{\mu}_{\text{IPW}}^{(b)}$. Define the bootstrap variance for $\mu$ by $s^2_{\text{boot}} = B^{-1} \sum_{b=1}^{B} (\tilde{\mu}_{\text{IPW}}^{(b)} - \tilde{\mu}_{\text{IPW}})^2$, where $\tilde{\mu}_{\text{IPW}} = B^{-1} \sum_{b=1}^{B} \tilde{\mu}_{\text{IPW}}^{(b)}$ and a bootstrap two-tailed 95% CI is constructed by $(\tilde{\mu}_{\text{IPW}} + q(0.025)s_{\text{boot}}, \tilde{\mu}_{\text{IPW}} + q(0.975)s_{\text{boot}})$, where $q(0.025)$ and $q(0.975)$ are the 2.5 and 97.5 percentiles of the standardized bootstrap samples of $(\tilde{\mu}_{\text{IPW}} - \tilde{\mu}_{\text{IPW}})/s_{\text{boot}}$ (see Theorem 23.5 of Van der Vaart 2000). The bootstrap CIs of $\tau^2$ based on $\tilde{\tau}^2_{\text{IPW}}$ are constructed in a similar manner.

### 3.5 Other measures for between-study heterogeneity

Higgins and Thompson (2002) discussed several heterogeneity measures alternative to $\tau^2$, including $H^2 = Q/(N - 1)$ and $I^2 = (H^2 - 1)/H^2$. The former can be interpreted approximately as the ratio of CI widths for the overall mean from random-effects and fixed-effect models, and the latter can be used to describe the percentage of variability for $\mu$ that is due to heterogeneity rather than sampling error.
The \( I^2 \) has been adopted by the Cochrane Collaboration as the summary measure of heterogeneity in their Review Manager Software and other commonly used packages for meta-analysis (e.g., metafor package, meta package). With the IPW version of Q-statistics (\( Q_{IPW} \)), the IPW versions of \( H^2 \) and \( I^2 \) can be defined as \( H^2_{IPW} = Q_{IPW} / (S - 1) \) and \( I^2_{IPW} = (H^2_{IPW} - 1) / H^2_{IPW} \), which would be useful to describe heterogeneity in the presence of the selective publication process.

4 | SIMULATION STUDY

4.1 | Settings

Simulation studies were carried out to assess the performance of the proposed IPW estimator. We conducted two types of simulation studies; one was based on one-parameter selection functions (3) and (4) and the other on two-parameter ones (5) and (6).

We begin with describing how to generate complete data of published and unpublished studies. The simulation design for generating all the studies was similar to that considered in Huang et al. (2021). Suppose we are interested in conducting a meta-analysis of randomized clinical trials to compare two treatment groups with a dichotomous outcome. The log-odds ratio was used as the summary measure of the treatment effect between the experimental group and the control group. We set the overall treatment effect \( \mu = -0.50 \), which was motivated by the Clopidogrel study in Web Appendix E, and \( \tau^2 = 0.0025 - 0.09 \), which reflected the small to moderate heterogeneity. The total number of studies including published and unpublished was set as 15, 25, 50, or 100. At first, we generated the true log-odds ratio of the \( i \)th study \( \mu_i \) from \( N(\mu, \tau^2) \). Next, we generated the true event rate in the control group \( p_{ic} \) from the uniform distribution \( U(0.2, 0.9) \) and then the event rate in the treatment group \( p_{it} \) could be derived as \( e^{\beta \mu_i} p_{ic} / (1 - p_{ic} + e^{\beta \mu_i}) \). Following Kuss (2015), the total sample size of each study was generated from \( LN(5, 1) \), the log-normal distribution with the location parameter 5 and scale parameter 1, and the minimum sample size was restricted to 20 patients (values below 20 were rounded up to 20). Subjects were allocated to the two treatment groups with probability of 0.5. Then, the individual participant data could be generated from the binomial distributions \( B(n_{ic}, p_{ic}) \) and \( B(n_{it}, p_{it}) \), respectively. With the generated individual participant data, we calculated the empirical log-odds ratio \( y_i \) and its standard error \( \sigma_i \).

From the complete data generated following the above-mentioned procedure, we selectively picked several studies according to one- or two-parameter selection models and then created four datasets, which are referred to as sDatasets 1 to 4, among which the first two were based on one-parameter selection functions and the latter two were based on two-parameter ones. The indicator of publication status \( D_i \) was generated from the binomial distribution \( B(1, \pi_i(\beta)) \). For sDataset 1, we selected published studies using the one-parameter logistic selection function (3) with \( \beta = 2 \). For sDataset 2, the one-parameter modified logistic selection function (4) with \( \beta = 5 \) was used. In these datasets, around 20% studies were regarded as unpublished. For sDataset 3 and sDataset 4, the two-parameter selection functions of Equations (5) and (6) with \( \beta = (-0.3, -1) \) were used, and about 25% studies in sDataset 3 and 30% studies in sDataset 4 were regarded as unpublished, respectively. The selection functions used to generate sDataset 3 and sDataset 4 are shown in Figure 1. For each dataset, 1000 meta-analyses were simulated.

4.2 | Results

Simulation results for one-parameter selection functions (sDatasets 1 and 2) are presented in Web Appendix D1. We observed that our IPW estimators outperformed the conditional likelihood-based methods in general with the moderate selection. In this section, we present the results for the simulation data generated with the two-parameter probit selection function (sDataset 3). Results for the two-parameter logistic selection function (sDataset 4) are provided in Web Appendix D (see also Section 4.3). In Tables 1 and 2, we used the column “Selection” to denote the selection function applied in estimation, in which “2-probit” referred to the two-parameter probit selection function and the “2-logit” referred to the two-parameter logistic selection function. The column “Status” was used to denote whether the estimate was from the correctly specified selection function or not; “C” means correctly specified and “M” means misspecified. We compared our proposed method with the maximum conditional likelihood method proposed by Copas (2013). As mentioned in Section 3.2, this method is implemented with a marginal selection probability fixed (sensitivity analysis). In order to make a fair comparison, we used the empirical publication rate \( (p = N/S) \) in implementation of the Copas method, and nlminb() function was used for its conditional log-likelihood optimization.

In Table 1, we present the simulation results of \( \mu \) estimates. For reference, we also show results with the standard mixed-effects model. The crude estimates with the standard mixed-effects model were highly biased suggesting that the simulation design successfully generated data under selective publication. Both the Copas sensitivity analysis method and the proposed IPW method could
TABLE 1  Simulation results for estimation of $\mu$ under two-parameter probit selection model (*Dataset 3*) with $\beta = (-0.3, -1.0)$ and $\tau = 0.05, 0.15, \text{ or } 0.30$

| $\tau^2$ | Method | Selection | Status | AVE(SD) | CP | LOCI | NOC | AVE(SD) | CP | LOCI | NOC | AVE(SD) | CP | LOCI | NOC | AVE(SD) | CP | LOCI | NOC |
|----------|--------|-----------|--------|---------|-----|------|-----|---------|-----|------|-----|---------|-----|------|-----|---------|-----|------|-----|
| 0.0025   | DL     |           |        | -0.552 (0.083) | 0.944 | 0.346 | 1000 | -0.547 (0.064) | 0.910 | 0.256 | 1000 | -0.545 (0.044) | 0.858 | 0.176 | 1000 | -0.543 (0.030) | 0.726 | 0.121 | 1000 |
|          | Copas  | 2-probit | C      | -0.511 (0.091) | 0.539 | 0.187 | 958  | -0.503 (0.073) | 0.563 | 0.154 | 990  | -0.498 (0.049) | 0.634 | 0.117 | 996  | -0.498 (0.034) | 0.673 | 0.085 | 999  |
|          | IPW (Asym) | 2-logit | M      | -0.524 (0.081) | 0.933 | 0.885 | 1000 | -0.520 (0.063) | 0.930 | 1.787 | 1000 | -0.518 (0.043) | 0.951 | 0.484 | 1000 | -0.519 (0.030) | 0.937 | 0.188 | 1000 |
|          | IPW (Boot) | 2-logit | M      | -0.524 (0.081) | 0.971 | 0.361 | 1000 | -0.519 (0.063) | 0.961 | 0.269 | 1000 | -0.518 (0.043) | 0.954 | 0.192 | 1000 | -0.519 (0.030) | 0.949 | 0.139 | 1000 |
|          | IPW (Asym) | 2-probit | C      | -0.508 (0.086) | 0.918 | 0.751 | 1000 | -0.502 (0.068) | 0.928 | 0.614 | 1000 | -0.497 (0.049) | 0.960 | 0.489 | 1000 | -0.497 (0.034) | 0.970 | 1.178 | 1000 |
|          | IPW (Boot) | 2-probit | C      | -0.508 (0.086) | 0.974 | 0.391 | 1000 | -0.502 (0.068) | 0.966 | 0.296 | 1000 | -0.497 (0.049) | 0.972 | 0.221 | 1000 | -0.497 (0.034) | 0.987 | 0.176 | 1000 |
|          | Pub [Q1,Q3] |        |        | 12.0 [11.0, 13.0] | 20.0 [18.0, 21.0] | - | - | - | 39.0 [37.0, 41.0] | - | - | - | 78.0 [75.0, 81.0] | - | - | - |
| 0.0225   | DL     |           |        | -0.569 (0.098) | 0.879 | 0.369 | 1000 | -0.565 (0.078) | 0.837 | 0.280 | 1000 | -0.561 (0.052) | 0.734 | 0.190 | 1000 | -0.560 (0.036) | 0.582 | 0.132 | 1000 |
|          | Copas  | 2-probit | C      | -0.522 (0.113) | 0.475 | 0.184 | 947  | -0.509 (0.093) | 0.526 | 0.158 | 990  | -0.506 (0.063) | 0.539 | 0.115 | 998  | -0.503 (0.043) | 0.605 | 0.087 | 999  |
|          | IPW (Asym) | 2-logit | M      | -0.532 (0.096) | 0.930 | 1.043 | 1000 | -0.526 (0.077) | 0.923 | 30.727 | 1000 | -0.524 (0.053) | 0.945 | 0.919 | 1000 | -0.525 (0.036) | 0.942 | 0.319 | 1000 |
|          | IPW (Boot) | 2-logit | M      | -0.532 (0.096) | 0.940 | 0.381 | 1000 | -0.526 (0.077) | 0.921 | 0.294 | 1000 | -0.524 (0.053) | 0.927 | 0.208 | 1000 | -0.525 (0.036) | 0.918 | 0.152 | 1000 |
|          | IPW (Asym) | 2-probit | C      | -0.514 (0.100) | 0.910 | 1.154 | 1000 | -0.503 (0.082) | 0.932 | 1.000 | 1000 | -0.500 (0.059) | 0.957 | 0.713 | 1000 | -0.498 (0.042) | 0.974 | 0.577 | 1000 |
|          | IPW (Boot) | 2-probit | C      | -0.514 (0.100) | 0.952 | 0.417 | 1000 | -0.503 (0.082) | 0.946 | 0.334 | 1000 | -0.500 (0.059) | 0.954 | 0.246 | 1000 | -0.498 (0.042) | 0.972 | 0.194 | 1000 |
|          | Pub [Q1,Q3] |        |        | 12.0 [10.0, 13.0] | 19.0 [18.0, 21.0] | - | - | - | 38.0 [36.0, 40.0] | - | - | - | 77.0 [74.0, 80.0] | - | - | - |

(Continues)
| $\tau^2$ | Method     | Selection | Status | $S = 15$ |          | $S = 25$ |          | $S = 50$ |          | $S = 100$ |          |
|---------|------------|-----------|--------|----------|----------|----------|----------|----------|----------|----------|----------|
|         |            |           |        | AVE(SD)  | CP       | LOCI     | NOC      | AVE(SD)  | CP       | LOCI     | NOC      | AVE(SD)  | CP       | LOCI     | NOC      | AVE(SD)  | CP       | LOCI     | NOC      |
| 0.0900  | DL         |           |        | −0.627 (0.126) | 0.790   | 0.452    | 1000    | −0.621 (0.096) | 0.712   | 0.341    | 1000    | −0.623 (0.067) | 0.510   | 0.246    | 1000    | −0.620 (0.046) | 0.231   | 0.176    | 1000    |
|         | Copas      | 2-probit C|        | −0.564 (0.154) | 0.373   | 0.197    | 973     | −0.551 (0.121) | 0.396   | 0.167    | 983     | −0.534 (0.094) | 0.392   | 0.118    | 999     | −0.518 (0.067) | 0.440   | 0.093    | 1000    |
|         | IPW        | 2-logit  M|        | −0.560 (0.126) | 0.892   | 1.700    | 1000    | −0.552 (0.093) | 0.936   | 1.219    | 1000    | −0.551 (0.067) | 0.942   | 1.555    | 1000    | −0.549 (0.047) | 0.935   | 1.030    | 1000    |
|         | IPW        | 2-logit  M|        | −0.560 (0.126) | 0.893   | 0.462    | 1000    | −0.552 (0.093) | 0.908   | 0.354    | 1000    | −0.551 (0.067) | 0.879   | 0.261    | 1000    | −0.549 (0.047) | 0.851   | 0.198    | 1000    |
|         | IPW        | 2-probit C|        | −0.538 (0.131) | 0.883   | 1.292    | 1000    | −0.523 (0.101) | 0.935   | 2.367    | 1000    | −0.514 (0.075) | 0.966   | 1.716    | 1000    | −0.507 (0.055) | 0.974   | 1.910    | 1000    |
|         | IPW        | 2-probit C|        | −0.538 (0.131) | 0.913   | 0.506    | 1000    | −0.523 (0.101) | 0.951   | 0.403    | 1000    | −0.514 (0.075) | 0.956   | 0.309    | 1000    | −0.507 (0.055) | 0.978   | 0.250    | 1000    |
|         | Pub        | [Q1,Q3]   |        | 11.0 [10.0] | −      | −      |        | 18.0 [17.0] | −      | −      |        | 37.0 [35.0] | −      | −      |        | 74.0 [71.0] | −      | −      |
|         | Pub        | [Q1,Q3]   |        | , 12.0 | −      | −      |        | , 20.0 | −      | −      |        | , 39.0 | −      | −      |        | , 77.0 | −      | −      |
|         | True       |           |        | −0.500 | −      | −      |        | −0.500 | −      | −      |        | −0.500 | −      | −      |        | −0.500 | −      | −      |

Note: Selection, the selection model used for estimation; 2-logit denotes the two-parameter logistic selection model, 2-probit denotes the two-parameter probit selection model; Status, model specification: C means selection model was correctly specified, M means selection model was misspecified; $s$, number of total studies; AVE, mean value of estimates; SD, standard error of estimates; CP, 95% confidence interval coverage probability; LOCI, length of confidence interval; NOC, number of converged cases in 1,000 realizations; DL, random-effects model with DerSimonian–Laird method; Copas, Copas’ sensitivity analysis method; IPW (Asym), the proposed method with asymptotic variance; IPW (Boot), the proposed method with parametric bootstrap confidence interval; Pub [Q1,Q3], median and 1st and 3rd quantiles of the number of published studies.
| $\tau^2$ | Method | Selection Status | $S = 15$ | $S = 25$ | $S = 50$ | $S = 100$ |
|---|---|---|---|---|---|---|
| 0.0025 | DL | | 0.007 (0.018) | 0.937 (0.100) | 0.005 (0.011) | 0.925 (0.010) | 0.003 (0.006) | 0.910 (0.048) | 0.001 (0.003) | 0.842 (0.021) | 0.826 |
| | IPW (Asym) 2-logit | M | 1.000 (0.100) | 740 (0.005) | 1.000 (0.011) | 732 (0.003) | 1.000 (0.007) | 771 (0.001) | 0.001 (0.004) | 1.000 (0.032) | 825 |
| | IPW (Boot) 2-logit | M | 1.000 (0.112) | 740 (0.005) | 1.000 (0.011) | 732 (0.003) | 1.000 (0.0058) | 771 (0.001) | 0.001 (0.004) | 0.999 (0.044) | 825 |
| | IPW (Asym) 2-probit | C | 0.999 (0.023) | 710 (0.008) | 0.996 (0.017) | 660 (0.007) | 0.993 (0.017) | 620 (0.005) | 0.996 (0.011) | 0.996 (0.295) | 595 |
| | IPW (Boot) 2-probit | C | 0.998 (0.023) | 710 (0.008) | 1.000 (0.017) | 660 (0.007) | 0.989 (0.017) | 620 (0.005) | 0.994 (0.011) | 0.994 (0.070) | 595 |
| 0.0225 | DL | | 0.015 (0.027) | 0.934 (0.287) | 0.013 (0.020) | 0.927 (0.143) | 0.009 (0.017) | 0.866 (0.071) | 0.008 (0.010) | 0.784 (0.040) | 384 |
| | IPW (Asym) 2-logit | M | 0.014 (0.026) | 0.964 (0.130) | 0.012 (0.021) | 0.960 (2.102) | 0.010 (0.015) | 0.920 (0.122) | 0.008 (0.012) | 0.826 (0.060) | 423 |
| | IPW (Boot) 2-logit | M | 0.014 (0.026) | 1.000 (0.133) | 0.012 (0.021) | 1.000 (1.024) | 0.010 (0.015) | 1.000 (0.072) | 0.008 (0.012) | 0.999 (0.056) | 423 |
| | IPW (Asym) 2-probit | C | 0.017 (0.031) | 0.963 (0.186) | 0.017 (0.027) | 0.967 (0.211) | 0.016 (0.022) | 0.957 (0.163) | 0.018 (0.020) | 0.931 (0.157) | 202 |
| | IPW (Boot) 2-probit | C | 0.017 (0.031) | 1.000 (0.158) | 0.017 (0.027) | 1.000 (0.132) | 0.016 (0.022) | 0.997 (0.103) | 0.018 (0.020) | 0.984 (0.089) | 202 |
| 0.0900 | DL | | 0.047 (0.057) | 0.918 (0.458) | 0.044 (0.042) | 0.894 (0.241) | 0.045 (0.032) | 0.802 (0.143) | 0.046 (0.023) | 0.600 (0.091) | 7 |
| | IPW (Asym) 2-logit | M | 0.045 (0.056) | 0.611 (0.252) | 0.046 (0.044) | 0.648 (0.213) | 0.049 (0.036) | 0.662 (0.327) | 0.051 (0.027) | 0.649 (0.224) | 12 |
| | IPW (Boot) 2-logit | M | 0.045 (0.056) | 0.877 (0.223) | 0.046 (0.044) | 0.851 (0.179) | 0.049 (0.036) | 0.833 (0.144) | 0.051 (0.027) | 0.814 (0.18) | 12 |
| | IPW (Asym) 2-probit | C | 0.049 (0.060) | 0.633 (0.244) | 0.054 (0.050) | 0.697 (0.652) | 0.062 (0.044) | 0.790 (0.295) | 0.068 (0.035) | 0.835 (0.555) | 4 |
| | IPW (Boot) 2-probit | C | 0.049 (0.060) | 0.919 (0.253) | 0.054 (0.050) | 0.919 (0.214) | 0.062 (0.044) | 0.954 (0.180) | 0.068 (0.035) | 0.959 (0.148) | 4 |

Note: Selection, the selection model used for estimation; 2-logit denotes the two-parameter logistic selection model; 2-probit denotes the two-parameter probit selection model; Status, model specification: C means selection model correctly specified, M means selection model misspecified; s, number of total studies; AVE, mean value of estimates; SD, standard error of estimates; CP, 95% confidence interval coverage probability; LOCI, length of confidence interval; NOZ, number of 0 estimates in 1,000 realizations; DL, random-effects model with DerSimonian–Laird method; IPW (Asym), the proposed method with asymptotic variance; IPW (Boot), the proposed method with parametric bootstrap confidence interval.
reduce the bias and ours had smaller bias in almost all the scenarios when the selection model was correctly specified. We observed that the profile likelihood method in the Copas sensitivity analysis yielded substantially narrow CIs of inaccurate coverage probabilities. The CIs for the IPW estimator based on asymptotic normal approximation in Section 3.3 might be so wide due to the nearly singular hessian matrix in some cases. On the other hand, the CIs based on parametric bootstrap seemed more reasonable, and the coverage probabilities were close to the nominal level in almost all the scenarios. We also observed that the mis-specification of selection function could introduce considerable bias, although the mis-specified IPW estimators were still less biased than the standard mixed-effects model. Similar findings were also observed in the two-parameter logistic selection function (see Web Table 8 in Web Appendix D).

Next, we present the simulation results of $\tau^2$ estimates. We observed that our IPW version of DerSimonian-Laird $\tau^2_{IPW}$ estimator had smaller bias and fewer zero estimates than the $\tau^2_{DL}$ estimator in most scenarios. The coverage probabilities of the CIs based on the asymptotic normal approximation were unsatisfactory when the true $\tau^2$ was 0.09, and the parametric bootstrap CI performs well overall. We also observed that mis-specification of the selection function did not have much impact on the performance of $\tau^2_{IPW}$.

**4.3 Additional simulation studies for the performance of the proposed inverse probability weighting method**

To further investigate the performance of the proposed IPW method, additional simulation studies were conducted with the consideration of a severe selection (only 40% studies were published) as well as a simulation study assuming no treatment effect ($\mu=0$). All the results are presented in Web-Appendix D. Here, we briefly summarize the key findings. With severely selective publication, our method is less successful in eliminating publication bias completely. However, our IPW estimator had smaller biases in most scenarios and the bootstrap CIs for our IPW methods had empirical coverage probabilities close to the nominal level. In the simulation study assuming no treatment effect, our IPW estimator did not have considerable bias even with severely selective publication. We observed that the parametric bootstrap CIs had inflated empirical type I errors in some scenarios. This phenomenon seemed to be a common issue when utilizing the bootstrap techniques in meta-analysis. Similar observations had been reported even without publication bias (Viechtbauer, 2007).

**5 | DATA APPLICATION**

We illustrate our proposed method with the antidepressant study which aimed to evaluate the improvement in depression symptoms of 12 antidepressant drugs, and the outcome was measured as the standardized mean difference between the treatment group and placebo group. In this study, Turner et al. (2008) identified 73 registered randomized clinical trials from the FDA registry, among them 50 were published and 23 were unpublished, and the selective publication process was suggested by the nature of data that most of the published studies showed statistical significance while unpublished studies did not (see Turner et al. 2008 for more details). Since their focus was the meta-analysis of studies used for licensing, only the
FDA registry was used for study searching and hence both the effect size and standard error were available for all the studies (published and unpublished). Although this was not a typical situation of meta-analysis, we used this dataset for an illustrative purpose of our proposed method. Regarding the overall mean of all the 73 studies with the standard mixed-effect model as the “gold standard”, we compared the performance of our proposed method and other competitive methods empirically. The “gold standard” of DerSimonian–Laird estimate with all the 73 studies was 0.344 with a 95% CI of [0.300, 0.388], while the DerSimonian–Laird estimate only with the 50 published studies was 0.409 with a 95% CI of [0.366, 0.453], indicating that the underlying selective publication process might have considerable influence on estimation (see Table 3).

First, we summarize the results with the one-parameter selection functions. We applied the one-parameter logistic selection function (3) and its modified version (4), the $\hat{\beta}$ were estimated as 7.168 (95% asymptotic CI: [3.106, 11.231]; 95% bootstrap CI: (6.205, 8.852); where the asymptotic CI means the CI based on the asymptotic normal approximation) and 47.722 (95% asymptotic CI: [21.524, 73.920]; 95% bootstrap CI: [43.056, 57.404]) with Equations (3) and (4), respectively. The resulting estimates of $\mu$ as well as those conditional likelihood-based estimators were summarized in Table 3. Preston’s conditional likelihood-based method gave the estimates of 0.355 (95% CI: [0.296, 0.414]) and 0.357 (95% CI: [0.301, 0.414]) with the one-parameter logistic selection function (3) and its modified version (4), respectively. Our IPW method gave the more conservative estimates as 0.333 (95% asymptotic CI: [0.283, 0.383]; 95% bootstrap CI: [0.267, 0.399]) and 0.339 (95% CI: [0.287, 0.392]; 95% bootstrap CI: [0.256, 0.416]), accordingly.

Next, we demonstrate the results with the two-parameter probit (5) and logistic (6) selection functions. As we mentioned in the last paragraph, one benefit of this dataset is it included all the information for both published and unpublished studies, hence an empirical comparison could be performed by checking the estimation of $\hat{\beta}=(\hat{\beta}_0, \hat{\beta}_1)$ using standard maximum likelihood estimation (MLE) applied to all the 73 studies and our estimating equations (12) to the 50 published studies. For the two-parameter probit (5) selection function, the estimated selection functions were plotted with solid line and dashed line in Figure 2A for MLE and our method, respectively. Note that the positive effect size may indicate the effectiveness of the antidepressant drugs against placebo. Hence, the estimated publishing probability decreases dramatically as the $t$-statistic of each study decreases. For the estimation using MLE, $\hat{\beta}_0 = -2.151$ (95% CI: [−3.206, −1.223]) and $\hat{\beta}_1 = 1.488$ (95% CI: [0.979, 2.097]); as to our estimation simply using the sample sizes of unpublished studies, we obtained $\hat{\beta}_0 = −1.645$ (95% asymptotic CI: [−2.707, −0.584]; 95% bootstrap CI: [−2.403, −1.115]) and $\hat{\beta}_1 = 1.627$ (95% asymptotic CI: [−0.007, 3.260]; 95% bootstrap CI: [0.971, 2.056]). The bootstrap CIs were shorter than the asymptotic ones and seemed more relevant. Observations for the two-parameter logistic (6) selection function were similar to this (Figure 2B). We explained such observations in simulation studies, and we trust the bootstrap CIs more (see Web Appendix D). With both selection functions, the null hypothesis of $\hat{\beta}_1 = 0$ was statistically significant, successfully suggesting a selective publication process behind. For the results of $\mu$ estimates with two-parameter selection functions, we applied the Copas selection model.

### Table 3 Summary of the statistical analysis for publication bias evaluation of Antidepressant study

| Description | Data | Method | Selection | $\mu$ (95% CI) | $P$-value | $\tau^2$ (95% CI) | $I^2$ |
|-------------|------|--------|-----------|---------------|-----------|-----------------|------|
| No adjustment | Published & Unpublished | DL | – | 0.344 [0.300, 0.388] | <0.001 | 0.008 [0.000, 0.027] | 0.228 |
| Published | DL | – | 0.409 [0.366, 0.453] | <0.001 | 0.000 [0.000, 0.009] | 0.000 |
| One-parameter | Published | Preston | 1-logit | 0.355 [0.296, 0.414] | <0.001 | 0.000 [0.000, 0.006] | - |
| Published | Preston | 1-mlogit | 0.357 [0.301, 0.414] | <0.001 | 0.000 [0.000, 0.006] | - |
| Published & Registry | IPW (Asym) | 1-logit | 0.333 [0.284, 0.382] | <0.001 | 0.017 [0.006, 0.027] | 0.376 |
| Published & Registry | IPW (Boot) | 1-logit | 0.333 [0.268, 0.399] | – | 0.017 [0.000, 0.056] | 0.376 |
| Published & Registry | IPW (Asym) | 1-mlogit | 0.339 [0.290, 0.389] | <0.001 | 0.015 [0.003, 0.027] | 0.348 |
| Published & Registry | IPW (Boot) | 1-mlogit | 0.339 [0.256, 0.416] | – | 0.015 [0.000, 0.062] | 0.348 |
| Two-parameter | Published | Copas | 2-probit | 0.373 [0.356, 0.405] | – | 0.000 | - |
| Published & Registry | IPW (Asym) | 2-probit | 0.330 [0.287, 0.373] | <0.001 | 0.017 [0.006, 0.028] | 0.384 |
| Published & Registry | IPW (Boot) | 2-probit | 0.330 [0.216, 0.415] | – | 0.017 [0.000, 0.067] | 0.384 |
| Published & Registry | IPW (Asym) | 2-logit | 0.339 [0.296, 0.382] | <0.001 | 0.015 [0.004, 0.026] | 0.353 |
| Published & Registry | IPW (Boot) | 2-logit | 0.339 [0.262, 0.399] | – | 0.015 [0.000, 0.051] | 0.353 |

Note: Preston, Preston’s conditional likelihood method; Copas, Copas’ sensitivity analysis method; IPW (Asym), the proposed IPW method using asymptotic variance; IPW (Boot), the proposed IPW method using parametric bootstrap confidence interval; 1-logit, the one-parameter logistic selection model, 1-mlogit, the one-parameter modified logistic selection model; 2-probit, the two-parameter probit selection model; 2-logit, the two-parameter logistic selection model.
FIGURE 2 Plots of estimated selective publication processes for the antidepressant study: two-parameter probit model ($\Phi(-1.645 + 1.627t_i)$); two-parameter logistic model ($\frac{\exp(-2.706 + 2.290t_i)}{1 + \exp(-2.706 + 2.290t_i)}$)

with the marginal selection probability fixed at $p = 50/73$ and obtained the estimate of 0.373 with a very short 95% CI of [0.356, 0.405]. Our IPW method gave the estimates of 0.330 (95% asymptotic CI: [0.287, 0.373]; 95% bootstrap CI: [0.216, 0.415]) and 0.339 (95% asymptotic CI: [0.296, 0.382]; 95% bootstrap CI: [0.262, 0.399]) with the two-parameter probit (5) and logistic (6) selection function, respectively. It seemed that in this study all these methods successfully eliminate certain publication bias.

We also compared the estimation of heterogeneity with the methods in Section 3. We observed that all the methods only relying on published studies gave zero estimates, while the proposed IPW version of DerSimonian–Laird $\hat{I}^2_{IPW}$ estimator gave the non-zero estimates. With 73 studies (published and unpublished), the $I^2$ was 22.8%. On the other hand, with only published 50 studies, it was estimated as 0%, whereas the IPW version of $I^2$ ranged from 34.8% to 38.4% with different selection functions (see Table 3). We also applied our proposed method to another meta-analysis called the Clopidogrel study (Chen et al., 2013), which included fewer studies (see Web Appendix E).

6 | DISCUSSION

In this paper, we introduced the IPW method to address the publication bias issue in meta-analysis. In contrast to Matsuoka et al. (2007) and Mathur and VanderWeele (2020), we can avoid extensive calculations in sensitivity analyses. The simplicity and flexibility of the IPW estimator allows us to handle various $t$-type selection functions, and as shown in the simulation study (see details in Web Appendix D), it can result in an improvement in estimation of both overall effect size and heterogeneity relative to the conditional likelihood-based methods by Preston et al. (2004) and Copas (2013). In addition, the proposed confidence intervals based on a parametric bootstrap outperformed the confidence intervals based on asymptotic normal approximation with coverage probabilities close to the nominal level in most scenarios and had more reasonable lengths of CIs. Hence, the use of parametric bootstrapping is recommended in practice. On the other hand, we observed that the parametric bootstrap confidence intervals might not be satisfactory in terms of coverage probabilities with a small number of studies. In standard meta-analysis, there are several proposals for improvement including Follmann and Proschan (1999), Hartung and Knapp (2001a), and Hartung and Knapp (2001b) as well as some Bayesian methods, see, for example, Friede et al. (2017) and Günhan et al. (2020). In the maximum likelihood estimation with clinical trial registries, Huang et al. (2021) encountered this problem and observed that incorporating the adjustment by Hartung and Knapp (2001a) and Hartung and Knapp (2001b) could improve small-sample performance. Comprehensive investigation to improve the small-sample performance of the IPW estimator is warranted.

The publication bias issue has long been recognized as a kind of missing data problem. However, there is a notable difference between the publication bias issue and
the general missing data problem. In general missing data problems, such as drop-out in clinical trials, the whole study population is clearly understood. In other words, we know how many subjects are missing and some information such as baseline covariates are available for missing subjects. In the publication bias issue, it is hard to define a complete study population since we only observed published studies. Due to this reason, well-developed missing data methodologies such as the IPW method are hard to be used in this area directly and most of the methods for publication bias rely on the funnel-plot symmetry. After many years of development of clinical trial registries, prospective registration has been widely accepted by triallists, and searching on clinical trial registries plays an increasingly important role when performing systematic reviews. This allows us to identify unpublished studies and thereby provides an idea of the complete data. It gives us the opportunity to handle the publication bias issue as a general missing data problem.

In our view, clinical trial registries should play an important role to fill the gap between the publication bias issue and the general missing data problem. Our development of the IPW estimator as well as the maximum likelihood estimation by Huang et al. (2021) fits this perspective. These two methods used different types of selection functions and thus complement each other. With these methods, we can address robustness of the results of meta-analysis against different selective publication processes described by the Heckman-type and the t-type selection functions. As reported in the simulation study section, our IPW estimator might suffer from bias when the selection model is misspecified. Bias was observed regardless whether one- or two-parameter selection function was used. Furthermore, we could not identify any general tendency on what kind of misspecification was less serious. Since it is impossible to identify the true selective publication process in reality, a comprehensive sensitivity analysis with multiple selection functions would be useful and is always recommended in practice. Although it is advantageous to utilize clinical trial registries to address the publication bias issue, there are several practical issues to be addressed in future research, which are discussed in Web Appendix F.

ACKNOWLEDGMENTS
We thank the Associate Editor and Reviewers for their valuable suggestions to this paper. The second author’s research was supported by JST CREST Grant Number JPMJCR1763 and MEXT Project for Seismology toward Research Innovation with Data of Earthquake (STAR-E) Grant Number JPJ010217. The last author’s research was partly supported by Grant-in-Aid for Challenging Exploratory Research (16K12403) and for Scientific Research (16H06299, 18H03208) from the Ministry of Education, Science, Sports and Technology of Japan.

DATA AVAILABILITY STATEMENT
The data of the Clopidogrel study referenced in Section 5 in this paper are available in the Supporting information of this paper.

ORCID
Ao Huang https://orcid.org/0000-0002-3989-7658
Satoshi Hattori https://orcid.org/0000-0002-0200-1408

REFERENCES
Baudard, M., Yavchitz, A., Ravaud, P., Perrodau, E. & Boutron, I. (2017) Impact of searching clinical trial registries in systematic reviews of pharmaceutical treatments: methodological systematic review and reanalysis of meta-analyses. BMJ, 356, j448.
Borenstein, M., Hedges, L.V., Higgins, J.P. & Rothstein, H.R. (2010) A basic introduction to fixed-effect and random-effects models for meta-analysis. Research Synthesis Methods, 1(2), 97–111.
Carpenter, J.R., Schwarzer, G., Rücker, G. & Künstler, R. (2009) Empirical evaluation showed that the Copas selection model provided a useful summary in 80% of meta-analyses. Journal of Clinical Epidemiology, 62(6), 624–631.
Chen, Y., Zhang, Y., Tang, Y., Huang, X. & Xie, Y. (2013) High-maintenance-dose clopidogrel in patients undergoing percutaneous coronary intervention: a systematic review and meta-analysis. PloS ONE, 8(10), e78549.
Cochran, W.G. (1954) The combination of estimates from different experiments. Biometrics, 10(1), 101–129.
Copas, J. (1999) What works: selectivity models and meta-analysis. Journal of the Royal Statistical Society: Series A (Statistics in Society), 162(1), 95–109.
Copas, J. & Shi, J.Q. (2000) Meta-analysis, funnel plots and sensitivity analysis. Biostatistics, 1(3), 247–262.
Copas, J.B. (2013) A likelihood-based sensitivity analysis for publication bias in meta-analysis. Journal of the Royal Statistical Society: Series C (Applied Statistics), 62(1), 47–66.
DeAngelis, C.D., Drazen, J.M., Frizelle, F.A., Haug, C., Hoey, J., Horton, R., et al. (2005) Clinical trial registration: a statement from the international committee of medical journal editors. Archives of Dermatology, 141(1), 76–77.
DerSimonian, R. & Laird, N. (1986) Meta-analysis in clinical trials. Controlled Clinical Trials, 7(3), 177–188.
Duval, S. & Tweedie, R. (2000) Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics, 56(2), 455–463.
Egger, M., Smith, G.D., Schneider, M. & Minder, C. (1997) Bias in meta-analysis detected by a simple, graphical test. BMJ, 315(7109), 629–634.
Follmann, D.A. & Proschan, M.A. (1999) Valid inference in random effects meta-analysis. Biometrics, 55(3), 732–737.
Friede, T., Röver, C., Wandel, S. & Neuenschwander, B. (2017) Meta-analysis of few small studies in orphan diseases. Research Synthesis Methods, 8(1), 79–91.
Gopalakrishnan, S. & Ganeshkumar, P. (2013) Systematic reviews and meta-analysis: understanding the best evidence in primary
healthcare. *Journal of Family Medicine and Primary Care*, 2(1), 9–14.

Günhan, B.K., Röver, C. & Friede, T. (2020) Random-effects meta-analysis of few studies involving rare events. *Research Synthesis Methods*, 11(1), 74–90.

Hart, B., Lundh, A. & Bero, L. (2012) Effect of reporting bias on meta-analyses of drug trials: reanalysis of meta-analyses. *BMJ*, 344, d7202.

Hartung, J. & Knapp, G. (2001a) On tests of the overall treatment effect in meta-analysis with normally distributed responses. *Statistics in Medicine*, 20(12), 1771–1782.

Hartung, J. & Knapp, G. (2001b) A refined method for the meta-analysis of controlled clinical trials with binary outcome. *Statistics in Medicine*, 20(24), 658–673.

Hattori, S. & Zhou, X.H. (2018) Sensitivity analysis for publication bias in meta-analysis of diagnostic studies for a continuous biomarker. *Statistics in Medicine*, 37(3), 327–342.

Higgins, J.P. & Thompson, S.G. (2002) Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21(11), 1539–1558.

Huang, A., Komukai, S., Friede, T. & Hattori, S. (2021) Using clinical trial registries to inform Copas selection model for publication bias in meta-analysis. *Research Synthesis Methods*, 12(5), 658–673.

Kott, P.S. & Chang, T. (2010) Using calibration weighting to adjust for nonignorable unit nonresponse. *Journal of the American Statistical Association*, 105(491), 1265–1275.

Kuss, O. (2015) Statistical methods for meta-analyses including information from studies without any events—add nothing to nothing and succeed nevertheless. *Statistics in Medicine*, 34(7), 1097–1116.

Marks-Anglin, A., Luo, C., Piao, J., Gibbons, M. B.C., Schmid, C.H., Ning, J. & Chen, Y. (2022) Embrace: an em-based bias reduction approach through copas-model estimation for quantifying the evidence of selective publishing in network meta-analysis. *Biometrics*, 78(2), 754–765.

Mathur, M.B. & VanderWeele, T.J. (2020) Sensitivity analysis for publication bias in meta-analyses. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 69(5), 1091–1119.

Matsuoka, N., Hasegawa, C. & Hamada, C. (2007) A practical method adjusting for publication bias in meta-analysis based on p-value. *Japanese Journal of Biometrics*, 28(1), 19–36.

Mavridis, D., Sutton, A., Cipriiani, A. & Salanti, G. (2013) A fully Bayesian application of the Copas selection model for publication bias extended to network meta-analysis. *Statistics in Medicine*, 32(1), 51–66.

Miao, W. & Tchetgen Tchetgen, E.J. (2016) On varieties of doubly robust estimators under missingness not at random with a shadow variable. *Biometrika*, 103(2), 475–482.

Morikawa, K. & Kim, J.K. (2021) Semiparametric optimal estimation with nonignorable nonresponse data. *The Annals of Statistics*, 49(5), 2991–3014.

Peters, J.L., Sutton, A.J., Jones, D.R., Abrams, K.R. & Rushton, L. (2007) Performance of the trim and fill method in the presence of publication bias and between-study heterogeneity. *Statistics in Medicine*, 26(25), 4544–4562.

Piao, J., Liu, Y., Chen, Y. & Ning, J. (2019) Copas-like selection model to correct publication bias in systematic review of diagnostic test studies. *Statistical Methods in Medical Research*, 28(10–11), 2912–2923.

Preston, C., Ashby, D. & Smyth, R. (2004) Adjusting for publication bias: modelling the selection process. *Journal of Evaluation in Clinical Practice*, 10(2), 313–322.

Schwarzer, G., Carpenter, J. & Rücker, G. (2010) Empirical evaluation suggests Copas selection model preferable to trim-and-fill method for selection bias in meta-analysis. *Journal of Clinical Epidemiology*, 63(3), 282–288.

Terrin, N., Schmid, C.H., Lau, J. & Olkin, I. (2003) Adjusting for publication bias in the presence of heterogeneity. *Statistics in Medicine*, 22(13), 2113–2126.

Thornton, A. & Lee, P. (2000) Publication bias in meta-analysis: its causes and consequences. *Journal of Clinical Epidemiology*, 53(2), 207–216.

Turner, E.H., Matthews, A.M., Linardatos, E., Tell, R.A. & Rosenthal, R. (2008) Selective publication of antidepressant trials and its influence on apparent efficacy. *New England Journal of Medicine*, 358(3), 252–260.

Turner, R.M., Omar, R.Z., Yang, M., Goldstein, H. & Thompson, S.G. (2000) A multilevel model framework for meta-analysis of clinical trials with binary outcomes. *Statistics in Medicine*, 19(24), 3417–3432.

Van der Vaart, A.W. (2000) *Asymptotic statistics*. Cambridge University Press.

Viechtbauer, W. (2007) Confidence intervals for the amount of heterogeneity in meta-analysis. *Statistics in Medicine*, 26(1), 37–52.

**SUPPORTING INFORMATION**

Web Appendices are available with this paper at the Biometrics website on Wiley Online Library; a brief review of the conditional likelihood-based methods and mathematical proofs referenced in Section 3, the complete simulation study results referenced in Section 4, another example of small meta-analysis called the Clopidogrel study referenced in Section 5 and a discussion of practical issues referenced in Section 6. The R code for implementation of our methods has also been made available at the Biometrics website.

**How to cite this article:** Huang, A., Morikawa, K., Friede, T., & Hattori, S. (2023) Adjusting for publication bias in meta-analysis via inverse probability weighting using clinical trial registries. *Biometrics*, 79, 2089–2102. [https://doi.org/10.1111/biom.13822](https://doi.org/10.1111/biom.13822)