Meta Analysis with Fixed, Unknown, Study-specific Parameters

Brian Claggett, Minge Xie, and Lu Tian*

Abstract

Meta-analysis is a valuable tool for combining information from independent studies. However, most common meta-analysis techniques rely on distributional assumptions that are difficult, if not impossible, to verify. For instance, in the commonly used fixed-effects and random-effects models, we take for granted that the underlying study-level parameters are either exactly the same across individual studies or that they are realizations of a random sample from a population, often under a parametric distributional assumption. In this paper, we present a new framework for summarizing information obtained from multiple studies and make inference that is not dependent on any distributional assumption for the study-level parameters. Specifically, we assume the study-level parameters are unknown, fixed parameters and draw inferences about, for example, the quantiles of this set of parameters using study-specific summary statistics. This type of problem is known to be quite challenging in statistical inference (c.f., Hall & Miller (2010)). We utilize a novel resampling method via the confidence distributions of the study-level parameters to construct confidence intervals for the above quantiles. We justify the validity of the interval estimation procedure asymptotically and compare the new procedure with the standard bootstrapping method. We also illustrate our proposal with the data from a recent meta analysis of the treatment effect from an antioxidant on the prevention of contrast-induced nephropathy.

KEY WORDS: Bootstrap; Confidence distribution; Extrema; Meta analysis; Robust methods; Ties;

*Brian Claggett is Instructor of Medicine (Biostatistics), Department of Cardiology, Harvard Medical School, Boston, MA 02115 (E-mail: *bclaggett@partners.org*). Minge Xie is Professor of Statistics, Department of Statistics and Biostatistics, Rutgers University, Piscataway, NJ 08854 (E-mail: *mxie@stat.rutgers.edu*). Lu Tian is Associate Professor, Department of Health Research & Policy, Stanford University School of Medicine, Palo Alto, CA 94305 (E-mail: *lutian@stanford.edu*).

1. INTRODUCTION

Meta-analysis is a potentially powerful tool for combining information from multiple, independent studies for making inference, for example, about the treatment difference between two comparative groups. The use of meta-analysis methods has grown substantially in recent years, with over 2000 papers per year published in PubMed, as of 2006 (Sutton and Higgins, 2008). Among these approaches, the fixed effect and random-effects models (particularly the DerSimonian-Laird approach) are two of the most commonly used models in meta-analysis. In practice, however, it is difficult, if not impossible, to verify the fundamental assumptions of these two models. That is, one assumes either that the study-specific parameters of interest are constant across studies in a fixed-effect model or that these parameters are realizations of a random sample from a population with a parametric distribution. The standard goodness of fit test is not informative for validating these models.

In this article, we consider a very general framework in which we do not make any assumptions about the underlying unknown parameters, either as a common constant across studies or being a realization of a random sample from a proper continuous or discrete distribution. Specifically, suppose that there are a fixed number, K, of independent studies. We only assume that for any given sample size, the study-level parameters are fixed, unknown parameters, denoted by $\theta_1, \ldots, \theta_K$, any of which may or may not be equal to one another without restriction. In studying the asymptotic properties of the associated inference procedure, θ_k's are allowed to depend on the sample size N. Relevant inferential problems in meta-analysis can then be formulated in the form of making inferences for $\theta^{(q)}$ for either a specific or a few $0 < q < 1$, where $\theta^{(q)}$ is the $(100q)$th percentile of the set of parameters $\Theta = \{\theta_1, \ldots, \theta_K\}$. Our question in this article is how to make inference for $\theta^{(q)}$ via individual study-specific summary statistics.

Let n_k be the sample size for the kth study and $N = \sum_{k=1}^{K} n_k$ be the total sample size of all K studies. For simplicity, we assume $\lambda_k = n_k/N$ is stabilized away from 0 as $N \to \infty$, although the condition can be slightly relaxed; c.f., Xie et al. (2009); Hall & Miller (2010). We also assume that, from the kth study, $k = 1, \ldots, K$, there is a $\sqrt{N}$- consistent estimator for θ_k, say $\hat{\theta}_k$, with a standard error estimate s_k. Denote by $\Theta = \{\theta_1, \ldots, \theta_K\}$, $\hat{\Theta} = \{\hat{\theta}_1, \ldots, \hat{\theta}_K\}$ and also $F_k(t) = \lim_{N\to\infty} P\{(\hat{\theta}_k - \theta_k)/s_k \le t\}$. In practical applications, $F_k(t)$ is often the cumulative distribution function of the standard normal distribution, i.e., $\hat{\theta}_k$ can be approximated by $N(\theta_k, s_k^2)$ for large N. Our problem is how to utilize $\{\hat{\theta}_k, s_k\}, k = 1, \ldots, K$, to make inference about, for example, the aforementioned $\theta^{(q)}$. Note that, for $q \in (0,1), \theta^{(q)}$ is equivalent to $\theta^{(m)}$, the mth ordered value of Θ, with $m = \lfloor qK \rfloor + 1$.

When the $(100q)$th percentile is rather extreme, (i.e. q is close to 0 or 1), it is quite challenging to make inferences accurately about $\theta^{(q)}$ (Hall & Miller, 2010; Wandler & Hannig, 2012). In general, when several θ's are "clustered around" $\theta^{(q)}$, the inferential problem becomes non-trivial (Xie et al., 2009; Hall & Miller, 2010). In their study of "the problem of constructing confidence intervals or hypothesis tests for extrema of parameters, for example of $\max\{\theta_1, \ldots, \theta_K\}$," Hall & Miller (2010) stated that this type of problem is one of the "important problems where standard bootstrap estimators are not consistent, and where alternative approaches . . . also face significant challenges." The approach recommended by Hall & Miller (2010) for this problem, as well as a set of more general forms of extreme parameters, was to construct a conservative confidence interval by introducing a constant c_α to enlarge the usual confidence interval and use bootstrapping to estimate (tune) the constant c_α. Although the approach may be practical, it is conservative and fails to directly address the difficult problem of making inference on the extrema and other quantiles of the parameters. Hall & Miller (2010) pointed out that the difficulty for this type of problem is due to the unknown 'tie' and ‘near tie’ cases and demonstrated mathematically that it is not possible to estimate the limiting distribution of $\hat{\theta}^{(m)}$ consistently in the near ties case. Here, the near ties case can be interpreted as that, based on the current sample size, one or several ‘near tie’ parameters θ_k’s are too close to be distinguished from the target parameter $\theta^{(m)}$; c.f., Xie et al. (2009); Hall & Miller (2010). A precise definition of a near tie set and its interpretation is provided later in Section 2.

In this paper, using the concept of confidence distributions (Xie & Singh, 2013), we propose a new and simple resampling method to construct confidence interval estimators for $\theta^{(m)}$, regardless of the presence or absence of such ties or near ties. This new resampling method can be viewed as an extension of the well-studied and widely-used bootstrap method, but enjoys a more flexible interpretation and manipulation. In the proposed method, we avoid the difficult problem of estimating the limiting distribution of $\hat{\theta}^{(m)}$. Rather, we directly construct an asymptotic confidence distribution for $\theta^{(m)}$, which can lead to asymptotically proper inference for the ordered parameter $\theta^{(m)}$. The problem explored in this paper is more general than that of Xie et al. (2011), which proposed the combination of confidence distributions for the purpose of meta-analysis in the setting with a single parameter of interest, relying on an assumption of either fixed effects, random effects arising from a normal distribution, or of a single parameter shared by a majority of studies. The present setting requires none of these assumptions.

The rest of the paper is arranged as follows. In Section 2, we introduce and review the idea of confidence distributions as frequentist distributional estimators, along with connections to the related bootstrap estimators. In Section 3, we propose a general method for deriving an asymptotic confidence distribution for a particular $\theta^{(m)}$, which depends on the choice of weights employed, and examine three reasonable weighting schemes. We discuss the properties of weights which will guarantee appropriate asymptotic coverage, and show that only one of the weighting schemes satisfies the stated condition. In Section 4, we discuss a tuning procedure to empirically obtain unknown tuning constants for the proposed approach which takes advantage of key properties of confidence distribution in order to improve the finite-sample inference. In Section 5, we present simulation results showing that our proposed weighting scheme provides appropriate coverage in diverse settings. In Section 6, we illustrate our method using data from a recently published meta-analysis investigating the effect of an antioxidant on nephropathy. Overall, the development in the current paper simultaneously addresses two important problems: it develops a general inference framework for meta-analysis and also provides a solution for the well-established difficult problem of making inference for extrema of parameters.

2. REVIEW OF CONFIDENCE DISTRIBUTIONS AND CD-RANDOM VARIABLES

A confidence distribution (CD) is often referred to as a sample-dependent distribution function that can represent confidence intervals of all levels for a parameter of interest (see, e.g., Cox (1958); Efron (1993); and the review in Xie & Singh (2013)). Cox (2013) stated that the confidence distribution approach provides "simple and interpretable summaries of what can reasonably be learned from data (and an assumed model)". For example, consider a simple normal sample $\mathbf{x} = \{x_i, k = 1, \ldots, n\}$, where $x_i \sim N(\mu, 1)$. It is well known that a point estimate can be obtained by $\bar{x}_n = \sum_{k=1}^{n} x_i/n$, and an interval estimate (e.g., 95% CI) can be obtained by $(\bar{x}_n - 1.96/\sqrt{n}, \bar{x}_n + 1.96/\sqrt{n})$. When making inference based on confidence distributions, we use the distribution $N(\bar{x}_n, 1/n)$, or more formally, in its cumulative distribution function form $H(\mu) = \Phi(\sqrt{n}(\mu - \bar{x}_n))$, to estimate μ. It is clear that $H(\mu)$ depends on the sample $\mathbf{x}$, and $H(\mu)$ is a distribution function on the parameter space of μ when given the sample $\mathbf{x}$. It is also easy to show that $(H^{-1}(\alpha/2), H^{-1}(1-\alpha/2)) =$

$(\bar{x}_n + \Phi^{-1}(\alpha/2)/\sqrt{n}, \bar{x}_n + \Phi^{-1}(1-\alpha/2)/\sqrt{n})$ provides a level $(1-\alpha)100\%$ CI for μ, for every $0 < \alpha \leq 1$. Furthermore, the median (or mean) of the distribution estimator $N(\bar{x}_n, 1/n)$ provides a point estimator $\bar{x}_n$ for μ, and the tail mass $H(b) = \Phi(\sqrt{n}(b-\bar{x}_n))$ provides a p-value for the one-sided hypothesis test $K_0 : \mu \leq b$ versus $K_1 : \mu > b$. As such, the confidence distribution approach is a useful tool that can provide meaningful answers for all questions related to statistical inference. In the context under consideration in this article, we use an asymptotic confidence distribution (c.f., Singh et al. (2005), Definition 1.1; Schweder & Hjort (2002))

$$H_k(t) = 1 - F_k\left(\frac{\hat{\theta}_k - t}{s_k}\right),$$

to estimate θ_k, for each $k = 1, 2, \ldots, K$, where $F_k(t) = \lim_{N\to\infty} P\{(\hat{\theta}_k - \theta_k)/s_k \leq t\}$. Often, the central limit theorem applies, and we have $F_k(\cdot) = \Phi(\cdot)$ where $\Phi(\cdot)$ is the cumulative distribution of the standard normal distribution. In this case,

$$H_k(t) = \Phi\left(\frac{t - \hat{\theta}_k}{s_k}\right) \tag{1}$$

and we use the distribution $N(\hat{\theta}_k, s_k^2)$ to estimate θ_k, for each $k = 1, 2, \ldots, K$.

For the given study-level summary statistic $\{\hat{\theta}_k, s_k^2\}$, the asymptotic confidence distribution $H_k(\cdot)$ is a cumulative distribution function on the parameter space of θ_k. We can construct a random variable ξ_k such that $\xi_k \big| \hat{\theta}_k, s_k^2 \sim H_k(\cdot)$. This simulated ξ_k is called a *CD random variable* (c.f., Xie & Singh (2013) and the references therein). Considering $H_k(\cdot)$ in (1), we simulate ξ_k by $\xi_k \big| \hat{\theta}_k, s_k^2 \sim N(\hat{\theta}_k, s_k^2)$. It follows that, asymptotically, we have

$$\left.\frac{\xi_k - \hat{\theta}_k}{s_k}\right| \hat{\theta}_k \sim \left.\frac{\hat{\theta}_k - \theta_k}{s_k}\right| \theta_k \qquad (\text{both} \sim N(0,1)).$$

This statement is exactly the same as the key justification for bootstrap, with ξ_k in place of the bootstrap sample mean $\hat{\theta}_k^*$. Thus, a CD random variable ξ_k can be viewed as a model-based bootstrap estimator of θ_k. Indeed, Xie & Singh (2013) demonstrated under a very general setting that a CD random variable ξ is in essence the same as a bootstrap estimator or a simple linear transformation of a bootstrap estimator. This close connection between the CD random variable and a bootstrap estimator motivates a possible view of treating the concept of confidence distribution as an extension of a bootstrap distribution, albeit the confidence distribution concept is much broader.

In this article, we utilize the CD random variable and develop a new simulation mechanism to broaden the applications of the standard bootstrap procedures. Since a CD random variable is not limited solely to use as a bootstrap estimator, this freedom allows us to utilize ξ_k more liberally, which in turn allows us to develop more flexible statistical approaches and inference procedures.

3. AN INFERENCE METHOD BASED ON CONFIDENCE DISTRIBUTIONS

3.1 Proposed Methodology

For simplicity of notation and clarity of presentation, we illustrate our methodology in this and next section using normal confidence distributions $H_k(t) = \Phi((t - \hat{\theta}_k)/s_k)$ as defined in (1). As stated in Hannig & Xie (2012), it is often the case that the confidence distributions are asymptotically normal when summary statistics are used. Furthermore, our proposed development can be directly extended to the general form of $H_k(t) = 1 - F_k((\hat{\theta}_k - t)/s_k)$ with only minor modifications.

Denote by ξ_k the CD random variable corresponding to $H_k(t) = \Phi((t - \hat{\theta}_k)/s_k)$, i.e.,

$$\xi_k | \hat{\theta}_k, s_k^2 \sim N(\hat{\theta}_k, s_k^2), \quad \text{for } i = 1, \ldots, K. \tag{2}$$

Given a particular realized set of $\{\xi_k, k = 1, \ldots, K\}$ from each of the K studies and a set of weights $\{w_{k,(m)}, k = 1, \ldots, K\}$ to be elaborated later, we consider the construction of a weighted average of ξ_k's:

$$\xi^* = \sum_{k=1}^{K} w_{k,(m)} \xi_k \Big/ \sum_{k=1}^{K} w_{k,(m)} \tag{3}$$

for the purposes of making inference on $\theta^{(m)}$. In particular, we can easily simulate $\{\xi_k, k = 1, \ldots, K\}$ according to (2) and compute ξ^* according to (3). If we repeat this a large number of times, we can obtain a set of ξ^*'s, which may represent a set of realizations of CD-random variables from a confidence distribution for the parameter $\theta^{(m)}$. If this is indeed the case, we can report the mean/median/mode of the ξ^*'s as a point estimate of $\theta^{(m)}$, and the empirical $(\alpha/2)100\%$ and $(1-\alpha/2)100\%$ quantiles of the ξ^*'s as the level $(1-\alpha)100\%$ confidence interval for $\theta^{(m)}$.

This general procedure is very simple. Naturally, different choices of the weights $w_{k,(m)}$ lead to different procedures, and each procedure's resulting validity depends on the choice of its weights. In particular, we investigate in this paper the following potential choices of weights:

<u>Choice 1</u>:

$$w_{k,(m)}^{[1]} = \mathbf{1}\{\hat{\theta}_k = \hat{\theta}^{(m)}\},$$

where $\mathbf{1}\{\cdot\}$ is an indicator function and $\hat{\theta}^{(m)}$ is the mth smallest $\hat{\theta}_k$.

<u>Choice 2</u>:

$$w_{k,(m)}^{[2]} = \mathbf{1}\{\xi_k = \xi^{(m)}\},$$

where $\xi^{(m)}$ is the mth smallest ξ_k.

Weights $w_{k,(m)}^{[1]}$ and $w_{k,(m)}^{[2]}$ both represent intuitive ways of estimating and making inference on $\theta^{(m)}$. The use of $w_{k,(m)}^{[1]}$ is equivalent to using the confidence distribution (and resulting confidence interval) associated with the m^{th} ordered $\hat{\theta}$. It is essentially a naive bootstrap approach, in which we first identify the study associated with the m^{th} ordered $\hat{\theta}$, then based on this single study, make inference for $\theta^{(m)}$. The use of $w_{k,(m)}^{[2]}$ corresponds to the use of the distribution of the mth ordered ξ_k, and is therefore equivalent to the conventional bootstrap estimator of $\theta^{(m)}$, as discussed in Hall & Miller (2010). Despite these intuitively attractive qualities, we will show that both sets of weights may lead to undesirable properties, depending on the true nature of the data. Specifically, both weights do not consider the potential ties among θ_k's and fail to fully reflect the uncertainty concerning which study is truly associated with $\theta^{(m)}$, leading us to focus on the following new proposed weighting scheme:

Choice 3:

$$w_{k,(m)}^{[3]} = \mathcal{K}\big(\xi_k - \xi^{(m)}, b_L, b_R\big)$$

where $\mathcal{K}$ is a kernel function, and b_L, b_R represent the left-side and right-side kernel bandwidths. While different kernel shapes may result in different finite-sample performance, we henceforth assume a simple rectangular kernel, such that $\mathcal{K}\big(\xi_k - \xi^{(m)}, b_L, b_R\big) = \mathbf{1}\{-b_L \le (\xi_k - \xi^{(m)}) \le b_R\}$, and an empirical turning procedure is proposed later in Section 4 to select data-adaptive bandwidths that may help to stabilize small sample performance. Written this way, it is easy to see that $w_{k,(m)}^{[3]}$ represents a generalization of $w_{k,(m)}^{[2]}$ and reduces to the bootstrap estimator when $b_L = b_R \equiv 0$.

Given the appropriate kernel bandwidth, this third option, similar to equation (4.4) of Xie et al. (2011), can appropriately handle a variety of scenarios by better reflecting the uncertainty surrounding the identification of the studies associated with $\theta^{(m)}$, avoid biases by filtering out unrelated studies in the inference, and in many cases, offer narrower confidence intervals than those obtained via the other weighting schemes by combining information from all studies associated with $\theta^{(m)}$.

As theoretical motivation for the superiority of $w_{k,(m)}^{[3]}$ over the alternative weighting schemes, we provide later in Section 3.2 a sufficient condition for any given weighting scheme that allows for the use of ξ^* for asymptotically valid inference for $\theta^{(m)}$. Namely, $w_{k,(m)}$ must converge to a positive constant if θ_k belongs to the tie or near tie set of $\theta^{(m)}$, as defined below, and zero otherwise. We show that this requirement is not satisfied by $w^{[1]}_{k,(m)}$ or $w^{[2]}_{k,(m)}$ when there are ties or near ties, but is satisfied by $w^{[3]}_{k,(m)}$ when $(b_L, b_R) = O(N^{-\delta}), \delta \in (0, \frac{1}{2})$ in any situation, regardless of the presence or absence of ties or near ties.

Before presenting theoretical results in Section 3.2, let us end this subsection by formally defining the tie and near tie sets. The same definition has also been utilized in Xie et al. (2009); Hall & Miller (2010). In particular, we denote by

$$\Theta^{(m)}_{\mathcal{J}} = \{k : \theta_k = \theta^{(m)}, k = 1, \dots, K\}$$

the "tie set" of $\theta^{(m)}$, representing the set of all θ's which are equal to the parameter of interest. We also denote by

$$\Theta^{(m)}_{\mathcal{N}} = \{k : |\theta_k - \theta^{(m)}| = O(N^{-1/2}), k = 1, \dots, K\}$$

the "near tie set" of $\theta^{(m)}$. The interpretation of the "near tie" definition is that, based on current sample size n_k, a "near tie" parameter θ_k cannot be distinguished from the target parameter $\theta^{(m)}$. An equivalent expression is that, for any $k \in \Theta^{(m)}_{\mathcal{N}}$, $(\hat{\theta}_k - \hat{\theta}^{(m)}) - (\theta_k - \theta^{(m)}) \neq o_p(|\theta_k - \theta^{(m)}|)$, which means that the difference between θ_k and $\theta^{(m)}$ is not of greater order than the standard error of its estimator. This near tie definition uses the idea of "local asymptotics" (c.f., e.g., van deer Vaart (1998) Chapter 7 or Small (2010), Section 5.6), in which we study the local behavior around a fixed value of the target parameter through a sequence of root-N rated parameters. The local asymptotic technique can help measure the performance of an estimator in finer detail and ensure its performance in moderate sample sizes. Specifically in our setup, when we focus on $\theta^{(m)}$, we examine how many of the θ_k's are the same as (ties of) $\theta^{(m)}$, and how many of them are in its root-N local neighborhood. This root-N neighborhood enables us to investigate the impact of those studies with true parameters that are very close to (or the same as) $\theta^{(m)}$, which we cannot distinguish with the current sample size N. Effectively, we treat the parameters as fixed constants with respect to given sample sizes. Strictly speaking, we may denote θ_k by θ_{kN} to emphasize its dependence on N. See also Xie et al. (2009) who also provided a real data example, ranking VA hospitals across the US to motivate the near tie definition. Similarly, for example, in the high-dimensional and penalized regression literature, it is also often required that the dimensionality of the unknown non-zero regression coefficients grow at some rate of sample size n (see, e.g., a review article by Fan & Lv (2011)).

Throughout the paper, we assume the following separation condition:

$$d_m N^{1/2} \to \infty, \qquad [\mathrm{C}_{sp}]$$

where $d_m = \min_{k \notin \Theta_{\mathcal{N}}^{(m)}} \left|\theta_k - \theta^{(m)}\right|$ is the minimal distance between the θ_j's inside and outside the near tie set $\Theta_{\mathcal{N}}^{(m)}$. The separation condition [C_{sp}] allows that the separation d_m tends to zero but at a slower rate than $N^{-1/2}$. Condition [C_{sp}] is in fact much weaker than the conventional assumption involving ties or no ties. Briefly speaking, the condition requires that the θ's outside the tie/near tie set are not too close and can be distinguished from $\theta^{(m)}$ when N is large. As such, [C_{sp}] covers the tie, near tie, and no tie cases, each as special cases of the general condition. As a practical matter, it is often very reasonable to assume that the true parameters $\{\Theta\}$ (e.g., the treatment effect across several slightly different patient populations) are constant with respect to the sample size. In such a common situation, $d_m = \min_{\theta_k \neq \theta^{(m)}} |\theta_k - \theta^{(m)}| \geq$

$\min_{i \neq j} |\theta_i - \theta_j|$, which is typically a positive constant bounded away from zero and [C_{sp}] is automatically met. Empirically one may examine the K study-specific CDs: if the observed CDs can be grouped into few "clearly separated" clusters, then [C_{sp}] is likely to hold. Condition [C_{sp}] is therefore much weaker than those assumptions imposed in the conventional fixed-effects and random effects models, since we only assume in our setting that $\theta_1, \theta_2, \ldots, \theta_K$ are unknown parameters and that we have no information regarding which ones are inside or outside the tie set.

Throughout the paper, we assume that both $\Theta_{\mathcal{T}}^{(m)}$ and $\Theta_{\mathcal{N}}^{(m)}$ are completely unknown other than that they contain at least one member, $\theta^{(m)}$. Thus, without loss of generality, we can assume the number of studies in the tie set $|\Theta_{\mathcal{T}}^{(m)}| \geq 1$. The 'near tie' case is much broader than the tie case: $\Theta_{\mathcal{T}}^{(m)} \subseteq \Theta_{\mathcal{N}}^{(m)}$. Thus $|\Theta_{\mathcal{N}}^{(m)}| \geq |\Theta_{\mathcal{T}}^{(m)}| \geq 1$. We present next a set of theoretical results using the more general near tie setup. All results remain valid if $\Theta_{\mathcal{N}}^{(m)}$ is replaced by $\Theta_{\mathcal{T}}^{(m)}$.

3.2 Asymptotic theorem and properties of proposed weighing schemes

The following set of asymptotic results suggest that ξ^* may be used to make inference for $\theta^{(m)}$, if weights are chosen appropriately. A proof of the theorem is provided in Appendix.

THEOREM 3.1. *Suppose that we can prove that a set of weights possesses the following property:*

$$\lim_{N \to \infty} w_{k,(m)} = \begin{cases} c_k & \text{if } k \in \Theta_{\mathcal{N}}^{(m)}, \\ 0 & \text{if } k \notin \Theta_{\mathcal{N}}^{(m)}, \end{cases} \quad \text{for } k = 1, 2, \ldots, K \tag{4}$$

for some constants $c_k > 0$. Then, as $N \to \infty$, we have the following:

(i)

$$\sum_{k=1}^{K} w_{k,(m)}\hat{\theta}_k \Big/ \sum_{k=1}^{K} w_{k,(m)} = \theta^{(m)} + o_p(1) \quad \textit{and} \quad \sum_{k=1}^{K} w_{k,(m)}^2 s_k^2 \Big/ \Big\{ \sum_{k=1}^{K} w_{k,(m)} \Big\}^2 = \{s^{(m)}\}^2 + o_p(1),$$

where $\{s^{(m)}\}^2 = \sum_{k\in\Theta_{\mathcal{N}}^{(m)}} c_k^2 s_k^2 / \big\{ \sum_{k\in\Theta_{\mathcal{N}}^{(m)}} c_k \big\}^2$. *Furthermore,*

$$\frac{\xi^* - \sum_{k=1}^{K} w_{k,(m)}\hat{\theta}_k / \sum_{k=1}^{K} w_{k,(m)}}{\sqrt{\sum_{k=1}^{K} w_{k,(m)}^2 s_k^2 / \{\sum_{k=1}^{K} w_{k,(m)}\}^2}}\Bigg|\hat{\Theta} \sim \frac{\sum_{k=1}^{K} w_{k,(m)}\hat{\theta}_k / \sum_{k=1}^{K} w_{k,(m)} - \theta^{(m)}}{\sqrt{\sum_{k=1}^{K} w_{k,(m)}^2 s_k^2 / \{\sum_{k=1}^{K} w_{k,(m)}\}^2}}\Bigg|\Theta, \qquad (5)$$

both converging asymptotically to a N(0, 1) distribution.

(ii) Let

$$H_*(t) = P(\xi^* \leq t|\hat{\Theta}), \quad \textit{for any } t \in \Xi,$$

with Ξ *being the parameter space of* $\theta^{(m)}$. *When* $t = \theta^{(m)}$, *we have* $H_*(\theta^{(m)}) \to U(0,1)$, *in distribution; Thus, by Definition 1.1 of Singh et al. (2005),* $H_*(\theta)$ *is an asymptotic CD for* $\theta^{(m)}$.

Theorem 3.1 guarantees that $H_*(\theta) = Pr(\xi^* \leq \theta|\hat{\Theta})$ is an asymptotic confidence distribution for $\theta^{(m)}$ when our choice of weight satisfies the requirement (4). Based on $H_*(\theta)$, we can make asymptotically valid inference, including point estimation, confidence intervals, p-values, etc., for $\theta^{(m)}$; c.f., Xie & Singh (2013) and references therein. We therefore rely on ξ^* to provide valid inference for $\theta^{(m)}$ asymptotically.

While Theorem 3.1 outlines only a sufficient condition for the construction of a proper asymptotic confidence distribution, failure to meet this condition would imply that a proposed weighting scheme either fails to assign positive weight to a study in the true tie set or improperly assigns positive weight to a study not in the tie set with positive probability, properties which are intuitively unappealing and would result in either loss of efficiency or introduction of potential bias.

It remains to show whether any of the three sets of weight choices satisfy the requirement (4) and, if so, under which conditions. Since the asymptotic properties of each of the proposed weighted estimators depend on the true unknown values of Θ, we begin with the simplest setting of no ties and move on to the more complicated settings of ties and near ties, including the particularly difficult case in which the presence of such ties or near ties to $\theta^{(m)}$ cannot easily be determined.

The 'no tie' case is the case in which $|\Theta_{\mathcal{N}}^{(m)}| = |\Theta_{\mathcal{J}}^{(m)}| = 1$, referring to the case that $\Theta_{\mathcal{N}}^{(m)}$ and $\Theta_{\mathcal{J}}^{(m)}$ have only one element, $\theta^{(m)}$. There may or may not be ties among the remaining θ_k's, $k \notin \Theta_{\mathcal{N}}^{(m)}$, but this is irrelevant to the problem at hand in making inference for $\theta^{(m)}$.

Lemma 3.1 below states that, under the no tie condition, $|\Theta_{\mathcal{N}}^{(m)}| = |\Theta_{\mathcal{J}}^{(m)}| = 1$, all three choices of weights listed in Section 3 satisfy the condition in (4). A proof is given in the Appendix.

LEMMA 3.1 (ANY WEIGHT; NO TIE CASE). *Suppose that $|\Theta_{\mathcal{N}}^{(m)}| = |\Theta_{\mathcal{J}}^{(m)}| = 1$, $\hat{\Theta}$ is asymptotically normal, i.e, $\sqrt{N}(\hat{\Theta} - \Theta)$ approximately follows a normal distribution with mean zero as the sample size $N \to \infty$, and also Condition [C_{sp}] holds. For $s = 1, 2$, we have*

$$\lim_{N\to\infty} w_{k,(m)}^{[s]} = \begin{cases} 1 & \text{if } \theta_k = \theta^{(m)}, \\ 0 & \text{if } \theta_k \neq \theta^{(m)}, \end{cases} \qquad \text{for } i = 1, 2, \ldots, K. \tag{6}$$

Furthermore, if we use $w_{k,(m)}^{[3]}$ with $b_L, b_R \propto \tau_N$, where $\tau_N / d_m \to 0$, and $\tau_N\sqrt{N} \to \infty$, then (6) also holds for $w_{k,(m)}^{[3]}$.

In conjunction with Theorem 3.1, we can infer from the lemma that in the no tie case, we can implement the proposed approach using any of the three weighting schemes to make asymptotically valid inference for $\theta^{(m)}$. In fact, since (6) holds for all $s = 1, 2, 3$, it is easy to verify, following the proof of Theorem 3.1, that the inference based on these three different choices of weights are asymptotically equivalent. As a result, any advantages of one weighting scheme over another in this setting will depend on finite-sample performance, to be explored via simulation in Section 5.

The problem is much more complicated in the presence of ties (i.e., $|\Theta_{\mathcal{J}}^{(m)}| > 1$) or near ties (i.e., $|\Theta_{\mathcal{N}}^{(m)}| > 1$). In this case, the weights $w_{k,(m)}^{[1]}$ or $w_{k,(m)}^{[2]}$ for $k \in \Theta_{\mathcal{J}}^{(m)}$ or $\Theta_{\mathcal{N}}^{(m)}$ converge to random quantities, rather than constants c_k. We provide below a very simple example in a special case to illustrate the phenomenon.

Example 3.1 (Example showing that $w_{k,(m)}^{[1]}$ and $w_{k,(m)}^{[2]}$ do not necessarily satisfy conditions of Theorem 3.1). Without loss of generality, consider a very simple example with $K = 2$ and $\theta_1 \equiv \theta_2$. For $m = 1$, $\Theta_{\mathcal{J}}^{(m)} = \Theta_{\mathcal{N}}^{(m)} = \{1, 2\}$, but $w_{1,(m)}^{[1]} = 1 - w_{2,(m)}^{[1]} = \mathbf{1}\{\hat{\theta}_1 = \min(\hat{\theta}_1, \hat{\theta}_2)\}$ is a binary random variable that equals 1 with probability $P\{\hat{\theta}_1 \leq \hat{\theta}_2\} = 1 - P\{\hat{\theta}_2 \leq \hat{\theta}_1\} = 0.5$. Thus, both $w_{1,(m)}^{[1]}$ and $w_{2,(m)}^{[1]}$ are (dependent) Bernoulli random variables, each with $p = 0.5$, therefore violating (4). Similarly, for $m = 1$, the second choice of weights $w_{1,(m)}^{[2]} = 1 - w_{2,(m)}^{[2]} = \mathbf{1}\{\xi_1 = \min(\xi_1, \xi_2)\}$ is a binary random variable that equals 1 with probability $P\{\xi_1 \leq \xi_2\} = E\big[P\{\xi_1 \leq \xi_2 | \hat{\Theta}\}\big] = E\big[\Phi(\{\hat{\theta}_2 - \hat{\theta}_1\}/\{s_1^2 + s_2^2\}^{1/2})\big] = 0.5$. Again, both $w_{1,(m)}^{[2]}$ and $w_{2,(m)}^{[2]}$ are (dependent) Bernoulli random variables, each with $p = 0.5$, also violating (4).

In the above tie case, if θ_1 and θ_2 differ slightly, with $\theta_2 = \theta_1 + \delta/\sqrt{N}$, where $\delta = O(1)$ as $N \to \infty$, then the near tie definition applies, with $\Theta_{\mathcal{N}}^{(m)} = \{1, 2\}$. For simplicity, let us further assume that $s_1^2 = s_2^2 = a^2/N$ for a constant $a > 0$. It follows again that $w_{1,(m)}^{[1]} = 1 - w_{2,(m)}^{[1]} = \mathbf{1}\{\hat{\theta}_1 = \min(\hat{\theta}_1, \hat{\theta}_2)\}$ is a Bernoulli random variable but with probability $P\{\hat{\theta}_1 \leq \hat{\theta}_2\} = P\left\{\hat{\theta}_1 - \theta_1 \leq \hat{\theta}_2 - \theta_2 + \delta/\sqrt{N}\right\} \to P\{Z_1 \leq Z_2 + \delta/a\} \in (0, 1)$, where Z_1 and Z_2 are independent N(0,1) random variables. Similarly, for the second choice of weights, $w_{1,(m)}^{[2]} = 1 - w_{2,(m)}^{[2]} = \mathbf{1}\{\xi_1 = \min(\xi_1, \xi_2)\}$ is a binary random variable with $P\{\xi_1 \leq \xi_2\} = E\big[P\{\xi_1 \leq \xi_2 | \hat{\Theta}\}\big] = E\big[\Phi(\{\hat{\theta}_2 - \hat{\theta}_1\}/\{s_1^2 + s_2^2\}^{1/2})\big] \to E\big[\Phi(\{Z_2 - Z_1 - \delta/a\}/\sqrt{2})\big] \in (0, 1)$. Clearly, both weights 1 and weights 2 violate (4).

Indeed, the condition (4) is satisfied only when θ_1 and θ_2 are sufficiently separated and they are no longer near ties, i.e. $\delta = \sqrt{N}|\theta_2 - \theta_1| \to \infty$, which is exactly Condition [C_{sp}].

In the case of more than two ties with either $|\Theta_{\mathfrak{J}}^{(m)}| > 2$ or $|\Theta_{\mathfrak{N}}^{(m)}| > 2$, the weights $w_{k,(m)}^{[1]}$ or $w_{k,(m)}^{[2]}$ for $i \in \Theta_{\mathfrak{J}}^{(m)}$ or $\Theta_{\mathfrak{N}}^{(m)}$ still converge to random quantities, rather than constants. The patterns are similar to, but more complicated than, those discussed in the case of $|\Theta_{\mathfrak{J}}^{(m)}| = 2$ or $|\Theta_{\mathfrak{N}}^{(m)}| = 2$ in Example 3.1. Clearly, neither $w_{k,(m)}^{[1]}$ nor $w_{k,(m)}^{[2]}$ satisfies the requirement (4) in this case, so we can no longer ensure that the results from Theorem 3.1 are valid. In these cases, one consequence is that the resulting estimators are downwardly biased, with the bias increasing with $|\Theta_{\mathfrak{J}}^{(m)}|$ and $|\Theta_{\mathfrak{N}}^{(m)}|$. Consequently, the distribution of ξ^* is no longer a valid CD for making inference on $\theta^{(m)}$. Our simulation results indeed confirm that these two sets of weights perform poorly in situations with ties or near ties. Poor performance of the standard bootstrap procedure, which corresponds to the use of the second sets of weights $w_{k,(m)}^{[2]}$, was also reported by Hall and Miller (2010).

In contrast, if we use $w_{k,(m)}^{[3]}$ with $b_L, b_R \propto \tau_N$, where $\tau_N/d_m \to 0$ and $\tau_N\sqrt{N} \to \infty$, then we can show that (4) is satisfied. In fact, the following lemma shows that the requirement (4) is satisfied by $w_{k,(m)}^{[3]}$ in any case, regardless of whether or not any ties or near ties exist, and regardless of whether or not their existence can be determined from the data. The lemma also includes a result for a slightly modified $w_{k,(m)}^{[3]}$, $\widetilde{w}_{k,(m)}^{[3]} = w_{k,(m)}^{[3]}/(s_k^2 N) \propto w_{k,(m)}^{[3]}/s_k^2$. A proof can be found in the Appendix.

Lemma 3.2 (Weight $w_{k,(m)}^{[3]}$; Any case). *Suppose that Condition [C_{sp}] holds and we use $w_{k,(m)}^{[3]}$ with $b_L, b_R \propto \tau_N$, where $\tau_N/d_m \to 0$, and $\tau_N\sqrt{N} \to \infty$. For any*

$1 \leq |\Theta_{\mathcal{J}}^{(m)}| \leq |\Theta_{\mathcal{N}}^{(m)}| \leq K$, *we have*

$$\lim_{N\to\infty} w_{k,(m)}^{[3]} = \begin{cases} 1 & \textit{if } k \in \Theta_{\mathcal{N}}^{(m)}, \\ 0 & \textit{if } k \notin \Theta_{\mathcal{N}}^{(m)}, \end{cases} \quad \textit{and} \quad \lim_{N\to\infty} \widetilde{w}_{k,(m)}^{[3]} = \begin{cases} \lambda_k/\sigma_k^2 & \textit{if } k \in \Theta_{\mathcal{N}}^{(m)}, \\ 0 & \textit{if } k \notin \Theta_{\mathcal{N}}^{(m)}, \end{cases} \tag{7}$$

for $k = 1, 2, \ldots, K$. *Here,* $\sigma_k = \lim_{N\to\infty} s_k n_k^{1/2}$.

This lemma, together with Theorem 3.1, provides theoretical support for the use of the weighted sum of CD random variables ξ^* to make inference for $\theta^{(m)}$ in all cases, if either $w_{k,(m)}^{[3]}$ or $\widetilde{w}_{k,(m)}^{[3]}$ is used. From (7), only studies inside the tie and near tie set will be included for making inference and the studies outside the tie set are filtered out, asymptotically. Thus, making inference using the proposed method with $w_{k,(m)}^{[3]}$ is asymptotically equivalent to using the average of the $\hat{\theta}_k$ in the tie set (assuming we were to know the true tie set). When s_k's are heteroscedastic, the modified version $\widetilde{w}_{k,(m)}^{[3]}$ could be used to improve the efficiency and power of the inference under the heuristic rationale of giving greater weight to studies containing more information. It can be shown that $\widetilde{w}_{k,(m)}^{[3]}$ is equivalent to the asymptotically most efficient inverse variance weighting when the sets $\Theta_{\mathcal{N}}^{(m)}$ and $\Theta_{\mathcal{J}}^{(m)}$ are known a priori (Xie et al., 2011). In any case, as long as there is a separation between the studies not tied with $\theta^{(m)}$ and those tied with $\theta^{(m)}$ as quantified in Condition [C_{sp}], our proposal provides a class of approaches that can lead us to asymptotically correct inference. Further details will be discussed in the next section regarding the tuning of the kernel widths.

4. PROPOSED ALGORITHM FOR TUNING THE BANDWIDTH PARAMETERS

While we can guarantee that $w_{k,(m)}^{[3]}$ or $\tilde{w}_{k,(m)}^{[3]}$ will provide appropriate asymptotic inference as long as the tuning parameters (b_L, b_R) converge to 0 at the proper rate, it is important in practice to be able to select an appropriate value for the tuning parameters (b_L, b_R) to ensure good finite sample performance. Specifically, we decompose the bandwidth parameters by defining

$$b_L = \tau_N \cdot c_L \quad \text{and} \quad b_R = \tau_N \cdot c_R,$$

where $\tau_N = O(N^{-\delta})$, for a fixed $0 < \delta < 1/2$, and $c_L, c_R = O(1)$ are positive constants. In general, we may use $\tau_N = (s^{(m)})^{2\delta}$, where $s^{(m)}$ is the standard error associated with $\hat{\theta}^{(m)}$. Details for the construction of a scale-invariant version of τ_N are found in the Appendix.

The constants $(c_L,\ c_R)$ can potentially impact the performance of the proposed approach in finite sample situations. For instance, if we use very large values of $(c_L,\ c_R)$, the bandwidths $(b_L,\ b_R)$ can be very large and our inference will mimic a fixed-effects analysis, which is only reasonable under the assumption that $|\Theta_T^{(m)}| = K$. On the other hand, if we use very small values of $(c_L,\ c_R)$, the bandwidths $(b_L,\ b_R)$ can be very close to 0; thus the performance of our weights will be similar to $w_{k,(m)}^{[2]}$, which we have shown to be asymptotically valid only when $|\Theta_T^{(m)}| = 1$. It therefore seems reasonable that the tuning constants should be relatively large when ties are present and relatively small when no ties are present.

We propose to choose the appropriate paired constants $(c_L,\ c_R)$ via a procedure similar to a "double-bootstrap" algorithm. Specifically, we generate multiple replicate "new" data sets under an assumed set of "known" parameters (say Θ^*, which is obtained based on a shrinkage of $\hat{\Theta} = \{\hat{\theta}_1, \ldots, \hat{\theta}_K\}$). We then apply our proposed procedure to the generated "new" data sets and study their performance in covering the target parameter of the assumed "known" parameter set Θ^* across a range, via grid search, of possible $(c_L,\ c_R)$ pairs. The pair of $(c_L,\ c_R)$ associated with the best performance is then chosen for use with the actual observed data. More precise details are given below.

First, a set of presumed "true values" Θ^* is obtained by shrinking the observed vector $\hat{\Theta}$ towards its mean, with the degree of shrinkage based on the ratio of within study variation ($\sum_{k=1}^{K} s_k^2/K$) to the total variation in $\hat{\Theta}$ ($\sum_{k=1}^{K} \hat{\theta}_k^2/K - \{\sum_{k=1}^{K} \hat{\theta}_k/K\}^2$). This shrinkage is necessary, as $\hat{\Theta}$ typically has greater spread than Θ. For instance, in a fixed-effects scenario $\theta_1 \equiv \ldots \equiv \theta_K$ but $var(\hat{\Theta}) = \sum_{k=1}^{K} \hat{\theta}_k^2/K - \{\sum_{k=1}^{K} \hat{\theta}_k/K\}^2 > 0$ for the observed $\hat{\Theta}$. Further details regarding this shrinkage process are provided in the Appendix. The mth ordered θ^*, denoted $\theta^{*(m)}$, serves as the "true" target parameter for the purposes of the subsequent resampling procedure. For the given Θ^* and $\{\hat{s}_1^2 \ldots \hat{s}_K^2\}$ from the observed data, a new set of "observed" data $\hat{\Theta}^* = \{\hat{\theta}_1^*, \ldots \hat{\theta}_k^*\}$ is generated with $\hat{\theta}_k^* \sim N(\theta_k^*, \hat{s}_k^2)$ for each $k = 1, \ldots, K$. The corresponding CD random variables are $\xi_k^* \sim N(\hat{\theta}_k^*, \hat{s}_k^2)$. Following (3), we compute $\xi^{**} = \sum_{k=1}^{K} w_{k,(m)} \xi_k^* \Big/ \sum_{k=1}^{K} w_{k,(m)}$ with $w_{k,(m)} = \mathbf{1}\{-c_L \cdot \tau_N \le (\xi_k^* - \xi^{*(m)}) \le c_R \cdot \tau_N\}$ for a given pair of $(c_L,\ c_R)$. Repeating the above calculation R times for the given set of "observed" data $\hat{\Theta}^* = \{\hat{\theta}_1^*, \ldots \hat{\theta}_k^*\}$, we then compute $\mathbb{B}(c_L, c_R) = \hat{P}(\xi^{**} \le \theta^{*(m)}) = \frac{\sum_1^R \mathbf{1}\{\xi^{**} \le \theta^{*(m)}\}}{R}$ over the R realizations of ξ^{**}. This $\mathbb{B}(c_L, c_R)$ is an estimate of $H_{**}(\theta^{*(m)})$ for the given pair of (c_L, c_R), where $H_{**}(t) = P(\xi^{**} \le t \big| \hat{\Theta}^*)$ is the asymptotic CD for $\theta^{*(m)}$ defined in Theorem 3.1. Repeat this process with B new "observed" data sets to obtain ordered $\mathbb{B}_{(1)}(c_L, c_R), \ldots, \mathbb{B}_{(B)}(c_L, c_R)$ for each possible bandwidth pair (c_L, c_R). By the definition of confidence distribution, in order to ensure a proper coverage at all confidence levels (c.f., Xie & Singh (2013), Definition 1), $H_{**}(\theta^{*(m)})$ is asymptotically distributed as U(0,1) as a function of sample data. Thus, $\mathbb{B}_{(1)}(c_L, c_R), \ldots, \mathbb{B}_{(B)}(c_L, c_R)$ should behave as U(0,1) order statistics with $E(\mathbb{B}_{(b)}) = b/(B+1)$, if the bandwidth pair (c_L, c_R) is chosen correctly. Therefore, the loss function to be minimized in our procedure is $\mathbb{L}(c_L, c_R) = \frac{1}{B} \sum_{b=1}^{B} \left\{\mathbb{B}_{(b)}(c_L, c_R) - \frac{b}{B+1}\right\}^2$ and we choose the pair of (c_L, c_R) that minimizes $\mathbb{L}(c_L, c_R)$ by grid search. Sometimes, $\mathbb{L}(c_L, c_R)$ is approximately constant near its minimum over a certain region of (c_L, c_R). See Appendix for details regarding computationally efficient and stable tuning in this scenario.

5. SIMULATIONS

In order to demonstrate both small and large sample properties of our proposed estimator under different scenarios, we generate random data $X_{kj} \sim N(\theta_k, 1)$, with $\theta_k, i \in \{1, 2, \ldots, K\}, 1 \leq j \leq n_k$, taking different values according to the particular scenario: (1) Ties: $\theta_k \equiv 0$ for all i; (2) Uniform: $\theta_k = \frac{2i}{K+1} - 1$; and (3) Normal: $\theta_k = \Phi^{-1}(\frac{i}{K+1})$. For each scenario, we consider $K = 7, 11$, or 21, and we let the sample size from each study $n_k = 40$ or 4000. Using 500 simulated data sets for each setting, we show the coverage and median width of the nominal 95% confidence intervals.

We consider each of $w^{[1]}, w^{[2]}$, and $w^{[3]}$ as proposed in Section 2. The results are shown below. Because of the symmetric setups considered, the coverage and median interval width for any $\theta^{(k)}$ will be identical to that for $\theta^{(K+1-k)}$. We therefore only report results for the $5th$, $25th$, and $50th$ percentiles.

For our proposed method using kernel smoothing, the results shown use the tuning procedure described in the previous section with $R = 200$ random samples drawn from each study's confidence distribution and $B = 40$ bootstrap replications. Simulation results are shown below for $K = 7$ and 21. Simulation results corresponding to $K = 11$ show similar patterns and are available upon request.

Method 1 in Tables 1 and 2 are the naive bootstrap method corresponding to weight $w^{[1]}_{k,(m)}$, Method 2 is the regular bootstrap method corresponding to weight $w^{[2]}_{k,(m)}$, and Method 3 is our proposed kernel method corresponding to weight $w^{[3]}_{k,(m)}$. We first note that Method 1 will always return confidence intervals of equal or greater width than those returned by Method 2. Correspondingly, we find many settings in

Table 1: Simulation results with $K = 7$: 95% Confidence Intervals

			Method 1		**Method 2**		**Method 3**	
Scenario	n_k	Quan	Coverage	Width	Coverage	Width	Coverage	Width
1	40	5th	0.798	0.605	0.068	0.439	0.880	0.238
1	40	25th	0.976	0.597	0.516	0.350	0.938	0.245
1	40	50th	1.000	0.602	0.986	0.312	0.992	0.364
1	4000	5th	0.810	0.060	0.064	0.045	0.918	0.023
1	4000	25th	0.986	0.061	0.510	0.035	0.938	0.023
1	4000	50th	1.000	0.061	0.978	0.031	0.932	0.023
2	40	5th	0.956	0.600	0.956	0.539	0.940	0.584
2	40	25th	0.984	0.596	0.982	0.489	0.950	0.522
2	40	50th	0.966	0.601	0.974	0.473	0.930	0.533
2	4000	5th	0.952	0.060	0.952	0.060	0.952	0.060
2	4000	25th	0.938	0.061	0.938	0.061	0.938	0.061
2	4000	50th	0.960	0.061	0.960	0.061	0.960	0.061
3	40	5th	0.936	0.595	0.946	0.573	0.932	0.589
3	40	25th	0.964	0.599	0.972	0.539	0.942	0.573
3	40	50th	0.956	0.596	0.962	0.509	0.950	0.563
3	4000	5th	0.952	0.062	0.952	0.062	0.952	0.062
3	4000	25th	0.938	0.061	0.938	0.061	0.940	0.061
3	4000	50th	0.960	0.061	0.960	0.061	0.960	0.061

Table 2: Simulation results with $K = 21$: 95% Confidence Intervals

			Method 1		**Method 2**		**Method 3**	
Scenario	n_k	Quan	Coverage	Width	Coverage	Width	Coverage	Width
1	40	5th	0.818	0.602	0.000	0.294	0.898	0.138
1	40	25th	1.000	0.598	0.148	0.203	0.938	0.151
1	40	50th	1.000	0.598	0.982	0.184	0.990	0.224
1	4000	5th	0.866	0.061	0.000	0.029	0.934	0.013
1	4000	25th	1.000	0.060	0.144	0.020	0.934	0.013
1	4000	50th	1.000	0.061	0.988	0.019	0.934	0.013
2	40	5th	1.000	0.600	0.978	0.388	0.976	0.444
2	40	25th	1.000	0.596	0.984	0.325	0.944	0.372
2	40	50th	1.000	0.599	0.990	0.323	0.946	0.377
2	4000	5th	0.950	0.061	0.950	0.061	0.948	0.061
2	4000	25th	0.946	0.060	0.946	0.060	0.940	0.061
2	4000	50th	0.932	0.060	0.932	0.060	0.932	0.061
3	40	5th	0.970	0.601	0.966	0.486	0.926	0.548
3	40	25th	0.994	0.595	0.988	0.383	0.928	0.441
3	40	50th	0.998	0.605	0.988	0.359	0.966	0.421
3	4000	5th	0.950	0.061	0.950	0.061	0.950	0.061
3	4000	25th	0.946	0.060	0.946	0.060	0.948	0.061
3	4000	50th	0.932	0.060	0.932	0.060	0.932	0.061

which the coverage of Method 2 is far below the nominal level (e.g. the Ties setting). This result matches the report of poor performance of the regular bootstrap approach in Hall & Miller (2010) on extrema of parameters. In almost all of these settings (except the extreme quantiles in the Ties setting), Method 1 will provide appropriate, but conservative, confidence intervals. Our proposed Method 3, on the other hand, is shown to have appropriate coverage levels in all settings, as well as noticeably narrower confidence interval widths relative to Method 1 in nearly all cases (except in the cases when the θ_k's are well separated, in which case the interval lengths are the same for all methods). Relative to the regular bootstrap estimator (Method 2), the intervals from our proposed method are asymptotically narrower, in the ties setting, for the few cases in which the bootstrap estimator does provide appropriate coverage. Furthermore, the interval widths are similar (and asymptotically equal) to those from Method 2 in the uniform and normal settings.

6. EXAMPLE

To illustrate our proposed methodology, we use the data from 14 studies which assessed the effect of an antioxidant (acetylcysteine) in preventing contrast-induced nephropathy, a leading cause of acquired acute reduction in kidney function (Bagshaw & Ghali, 2004). The outcome of interest in each study was incidence of contrast-induced nephropathy, and so the parameter of interest was the odds ratio for the association between antioxidant usage and incidence of nephropathy. The summary data for each study is shown below.

A fixed effects analysis of this data by Bagshaw & Ghali (2004) resulted in a 95% confidence interval of (0.41, 0.87) for the (assumed) common odds ratio. However, significant heterogeneity was found in the study-level treatment effects (p=0.032). Thus, a random effects analysis was performed in Bagshaw & Ghali (2004), assuming that

Table 3: Summary results of 14 studies of acetylcysteine for prevention of contrast-induced nephropathy

Study	N	OR	95% CI
Allaqaband	85	1.23	(0.39, 3.89)
Baker	80	0.20	(0.04, 1.00)
Briguori	183	0.57	(0.20, 1.63)
Diaz-Sandova	54	0.11	(0.02, 0.54)
Durham	79	1.27	(0.45, 3.57)
Efrati	49	0.19	(0.01, 4.21)
Fung	91	1.37	(0.43, 4.32)
Goldenberg	80	1.30	(0.27, 6.21)
Kay	200	0.29	(0.09, 0.94)
Kefer	104	0.63	(0.10, 3.92)
MacNeill	43	0.11	(0.01, 0.97)
Oldemeyer	96	1.30	(0.28, 6.16)
Shyu	121	0.11	(0.02, 0.49)
Vallero	100	1.14	(0.27, 4.83)

the logs of the study-level odds ratios are normally distributed, which resulted in a somewhat wider confidence interval (0.32, 0.91).

Below we show the resulting 95% confidence intervals for each of the 14 ordered study-level treatment effects. The three columns of confidence intervals correspond to the weighting methods discussed in this article, with the third column representing our proposed procedure, which we have shown in simulations to have appropriate coverage, regardless of whether any or all of the true treatment effects are equal across studies. Given the heavy overlapping among resulting confidence intervals, the effect of ties/near ties cannot be ignored and thus the weights $w^{[1]}_{k,(m)}$ and $w^{[2]}_{k,(m)}$ should not be used. Even though we have some evidence to reject the fixed effects assumption, in this example it is particularly difficult, due to small sample sizes, to assess with any certainty whether or not any subsets of the study parameters are equal to one another, or whether the assumption of a normal distribution for the true study-specific log-odds-ratios used in the random effect model is justified.

We note that, in general, the intervals provided by Method 1 are essentially a re-ordering of the original study intervals, and thus do not provide substantially new information in terms of summarizing the treatment effects. The bootstrap intervals corresponding to Method 2 are noticeably narrower in some cases; however, it is alarming that the bootstrap interval for $\theta^{(14)}$, (1.44, 9.56), excludes even the maximum estimated treatment effect (estimated odds ratio = 1.37 from the Fung study). Using our proposed weights $w^{[3]}_{k,(m)}$ (Method 3) with the scale-invariant version of τ_N, we estimate that six of the fourteen studies exhibited significant treatment effects, while the remaining eight studies were found to be neutral. The confidence intervals for the 7th and 8th ordered treatment effects are (0.29, 1.26) and (0.31, 1.36), respectively. Using the conventional method of averaging the $(K/2)^{th}$ and $(K/2+1)^{th}$ ordered observations to estimate the median when K is an even number, we obtain a confidence interval of (0.30, 1.31) for the "median" treatment effect across these studies. This interval is slightly wider than the previously reported random effects analysis, though our inference is free of any distributional assumptions regarding the true values of the study-level treatment effects. Furthermore, if the true distribution of the parameters is not symmetric on the log scale, then our estimate of the median treatment effect will not necessarily be directly comparable to the random effects analysis, which estimates the mean of the random-effects distribution.

In Figure 1, we present the 95% confidence intervals for each ordered element of $\{\Theta\}$, with point estimates given by the mean of the associated CD. For comparison, the confidence intervals for the fixed-effects and random-effects meta-analysis are denoted by the vertical solid and dashed lines, respectively. Our estimates for $\theta^{(7)}$ and $\theta^{(8)}$ are highlighted for comparison. From Figure 1, we see that the six best performing trials suggest that acetylcysteine can prevent contrast-induced nephropathy, but we can not reach such a conclusion for the remaining eight trials.

Table 4: 95% Confidence Intervals for ordered study-level treatment effects (odds ratios) using nephropathy data

OS	CI (Method 1)	CI (Method 2)	CI (Method 3)
1	(0.02, 0.48)	(0.01, 0.13)	(0.03, 0.62)
2	(0.02, 0.51)	(0.03, 0.20)	(0.06, 0.59)
3	(0.01, 0.94)	(0.05, 0.28)	(0.06, 0.57)
4	(0.01, 4.64)	(0.07, 0.40)	(0.07, 0.59)
5	(0.04, 1.04)	(0.12, 0.54)	(0.12, 0.71)
6	(0.09, 0.94)	(0.17, 0.70)	(0.18, 0.97)
7	(0.19, 1.67)	(0.25, 0.91)	(0.29, 1.26)
8	(0.10, 3.85)	(0.33, 1.16)	(0.31, 1.36)
9	(0.27, 4.93)	(0.45, 1.46)	(0.43, 1.58)
10	(0.39, 3.94)	(0.56, 1.79)	(0.39, 1.52)
11	(0.45, 3.49)	(0.70, 2.27)	(0.46, 1.42)
12	(0.26, 6.06)	(0.87, 2.99)	(0.45, 1.62)
13	(0.28, 6.14)	(1.09, 4.38)	(0.39, 1.19)
14	(0.44, 4.26)	(1.44, 9.56)	(0.34, 1.20)

While our proposed procedure was motivated by a desire to avoid making any assumptions about the existence or nature of the distribution of our quantity of interest $\{\Theta\}$, we note that a plot such as that given in Figure 1 may resemble an empirical cumulative distribution function for the "true" distribution $F(\Theta)$. As sample size increases, the confidence distribution estimates for each $\theta^{(m)}$ converge to the true values $(\theta^{(1)}, \theta^{(2)}, ..., \theta^{(K)})$. If it can further be assumed that $(\theta^{(1)}, \theta^{(2)}, ..., \theta^{(K)})$ are a random sample from some overall distribution $F(\Theta)$ and also assumed K goes to infinity, then it can be seen that $\tilde{\theta}^{(q)} = \theta^{(\lfloor qK \rfloor + 1)}$ will converge, as K grows large, to $F_{\Theta}^{-1}(q)$.

In an attempt to assess the robustness of our procedure in a realistic setting in which the assumption of normality of study-level confidence distributions may not hold, we also performed a simulation study mimicking the data structure of the well-known rosiglitazone data set, previously analyzed in Tian et al. (2009). This data set features

48 randomized comparative studies of the diabetes drug rosiglitazone vs control, and we focus on the occurrences of myocardial infarction (MI) in each treatment arm. A key feature of the data is the low event rate (31 of the 48 trials featured ≤ 1 events), and thus large-sample approximations may not be valid. Tian et al. (2009), using binomial confidence intervals, assumed a constant risk difference in the event rates across studies and reported a 95% confidence interval of $(-0.08, 0.38)\%$ for the non-significantly increased risk associated with rosiglitazone. In our simulation study, we randomly generated 500 data sets, assuming the true event rates in each arm of each study is given by $(x+0.5)/(N+1)$, where (x, N) represent the observed number of MI's and total sample in a given study arm, respectively. We then applied our proposed procedure, sampling 200 times from the binomial CD for the risk difference in each study, omitting studies with sample sizes larger than 500 in order to focus on small-sample performance. Due to the discrete nature of the data, we omit the shrinkage step in the tuning procedure. We examined the $25th$, $50th$, and $75th$ percentiles of the study-specific parameters, and found that the 95% confidence intervals from our proposed method provided appropriate coverage for each percentile. Method 1 provided conservative coverage, with intervals approximately 2-3 times the width of those from our proposed method, and Method 2 was found to provide appropriate coverage only for the $50th$ percentile, but exhibited severe under-coverage for the $25th$ and $75th$ percentiles. These results are shown below in Table 5. When applied to the full Avandia data set analyzed by Tian et al. (2009), we report a 95% confidence interval of $(-0.07, 0.46)\%$ for the "median" treatment effect, with intervals of $(-0.27, 0.34)\%$ and $(0.06, 0.68)\%$ for the $25th$ and $75th$ percentiles, respectively.

Table 5: Simulation results using data mimicking rosiglitazone data from Tian et al. (2009)

Quantile	Ordered Risk Difference	95% Interval Coverage	Width
	Method 1		
25th	**8**	1.000	0.049
50th	**15**	1.000	0.040
75th	**22**	1.000	0.048
	Method 2		
25th	**8**	0.888	0.019
50th	**15**	0.986	0.012
75th	**22**	0.734	0.017
	Method 3		
25th	**8**	0.912	0.017
50th	**15**	0.976	0.012
75th	**22**	0.942	0.016

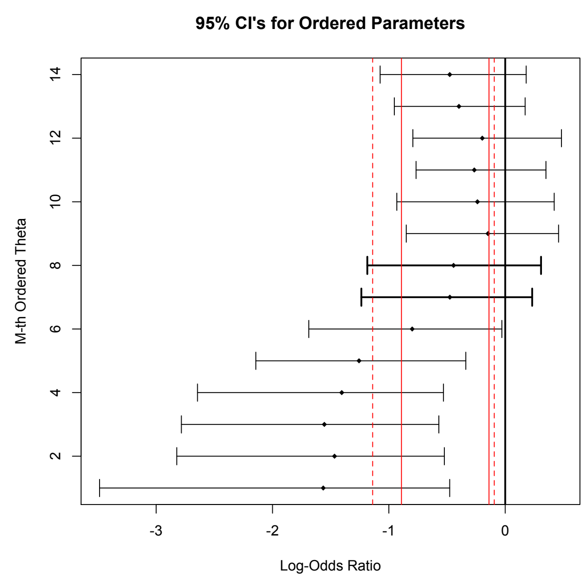

Figure 1: Confidence distribution estimates of treatment effects from 14 studies of acetylcysteine on nephropathy: Vertical solid (dashed) lines represent 95% CI from fixed-effects (random-effects) meta-analysis

7. DISCUSSION

In this paper, we introduce a simple and effective approach which simultaneously addresses two important problems. By introducing a procedure for making inference on *any* ordered value of a set of parameters, we may provide a summary of the treatment effects observed over a collection of studies without having to rely on any assumptions about the nature of or relationship between those treatment effects, thus enabling a non-parametric-like, model-free form of meta-analysis. Although we examine three different weighting schemes for inferential purposes, we find that only one, $w^{[3]}_{k,(m)}$, is appropriate in all settings, while the other two approaches are shown to be asymptotically equivalent, and appropriate, only if it is known a priori that no other components of $\{\Theta\}$ are equal to the parameter of interest. However, such knowledge is rarely available in practice. Therefore, while the resulting confidence interval from the proposed procedure might sometimes be wider than those provided by methods with more restrictive assumptions, the general applicability of our new method is appealing and may serve as a good point of comparison, just as many analysts now present results corresponding to both fixed-effects and random-effects meta-analysis models.

As with any meta-analysis, it is important to consider whether information from a variety of studies *should* be combined. Substantial heterogeneity between study-level parameters may be an indicator that the studies are in fact not comparable. The well known James-Stein estimator, for example, provides improved global estimation of multiple parameters without assuming any relationship between the parameters, though it is neither guaranteed to be optimal for estimating individual components nor provide a valid inference procedure.

It is quite possible that any departure from a fixed-effects assumption points to unexplained variability that needs to be further investigated. Traditionally the between-study variability is studied and accounted for via assumption of a normal distribution for the random effect. However, the normal distribution is often inadequate for describing complex variability. Indeed, oftentimes the limited number of studies would make us hesitate to make any parametric assumptions regarding the distribution of study-level effects. For example, suppose a meta-analysis involves only 7 studies. Even if we could know all the true study-level effects without error, it would require a great leap of faith to summarize the data by the estimated mean and variance, assuming that the 7 values are drawn from a normal distribution. On the other hand, a more practical and informative summary of 7 observations would simply be the values themselves, sorted from the smallest to the largest. Our proposal aims to do exactly this when the random effects can only be estimated with errors.

Additionally, our procedure also allows us to make inference on the extreme values of a set of parameters, a well-established problem that has proven to be intractable with respect to many statistical approaches. By taking advantage of the flexibility afforded by confidence distributions as functional estimators, as well as a tuning technique that accounts for the unknown presence or absence of ties and near-ties in small-sample settings, we are now able to provide valid inference in a wide variety of settings. Although our development has been presented under the setting of normal CDs, it can be extended to the setting of more general CDs. The observed good performance of the proposed method in a realistic small-sample setting with sparse data in Section 5 seems to also support the generality of the approach.

Lastly, we emphasize that in the current development, we assume that the number of studies K is finite, and reasonably small, and that each study sample size n_k goes to infinity, covering meta-analysis of a small number of large studies. In practice, it is also possible that we have a large number of trials that require meta-analysis. In this case with K going to infinity, we typically will have more information to help us to draw inference. For instance, in the standard normal random effects model, we can consistently estimate the underlying super-population parameter when we have a large number of studies (even if each study sample size is small). In addition, it is possible to investigate the same model-free meta-analysis problem when K goes to infinity. That is, we can make inference about $\theta^{(m)}$ without assuming that the underlying $\theta_1, \theta_2, \ldots, \theta_K$ are from a specific distribution. This inference problem is related to the classical compound decision approach (Copas (1969)). For a normal sample problem parallel to our investigation (but with K going to infinity and $n_k \equiv 1$), we refer readers to Jiang & Zhang (2009); Brown & Greenshtein (2009) who provided efficient empirical Bayesian approaches to estimate the unknown θ_k's and its empirical distribution. See, also, Zhang (2003) for a general review of the classical compound decision theory and empirical Bayes method.

APPENDIX

A1. Proof of Theorem 3.1.

(i) The first two results follow immediately from (4) and the fact that $|\hat{\theta}_k - \theta^{(m)}| \leq |\hat{\theta}_k - \theta_k| + |\theta_k - \theta^{(m)}| = O_p(N^{-1/2})$ for any $\theta_k \in \Theta_{\mathcal{N}}$. We only need to prove (5).

Note that, $\hat{\theta}_k \sim (\theta_k, s_k^2)$, for any i, it follows that

$$\frac{\sum_{k\in\Theta_{\mathcal{N}}} c_k\hat{\theta}_k / \sum_{k\in\Theta_{\mathcal{N}}} c_k - \sum_{k\in\Theta_{\mathcal{N}}} c_k\theta_k / \sum_{k\in\Theta_{\mathcal{N}}} c_k}{\sqrt{\sum_{k\in\Theta_{\mathcal{N}}} c_k^2 s_k^2 / \{\sum_{k\in\Theta_{\mathcal{N}}} c_k\}^2}} \sim N(0,1).$$

Again, from (4) and the fact that $|\theta_k - \theta^{(m)}| = O(N^{-1/2})$ for any $\theta_k \in \Theta_{\mathcal{N}}$, we have $\sum_{k=1}^K w_{k,(m)}\hat{\theta}_k = \sum_{k\in\Theta_{\mathcal{N}}} c_k\hat{\theta}_k + o_p(1)$, $\sum_{k=1}^K w_{k,(m)}^2 s_k^2 = \sum_{k\in\Theta_{\mathcal{N}}} c_k^2 s_k^2 + o_p(1)$, $\sum_{k=1}^K w_{k,(m)} = \sum_{k\in\Theta_{\mathcal{N}}} c_k + o_p(1)$ and $\sum_{k\in\Theta_{\mathcal{N}}} c_k\theta_k = \{\sum_{k\in\Theta_{\mathcal{N}}} c_k\}\,\theta^{(m)} + O(N^{-1/2})$. Thus, we have

$$\frac{\sum_{k=1}^K w_{k,(m)}\hat{\theta}_k / \sum_{k=1}^K w_{k,(m)} - \theta^{(m)}}{\sqrt{\sum_{k=1}^K w_{k,(m)}^2 s_k^2 / \{\sum_{k=1}^K w_{k,(m)}\}^2}} \to N(0,1), \quad \text{as } N \to \infty \qquad \text{(A.1)}$$

On the other hand, since $\xi^* = \sum_{k=1}^K w_{k,(m)}\xi_k / \sum_{k=1}^K w_{k,(m)}$ and ξ_k are CD random variables from $N(\hat{\theta}_k, s_k^2)$, we have

$$\left. \frac{\xi^* - \sum_{k=1}^K w_{k,(m)}\hat{\theta}_k / \sum_{k=1}^K w_{k,(m)}}{\sqrt{\sum_{k=1}^K w_{k,(m)}^2 s_k^2 / \{\sum_{k=1}^K w_{k,(m)}\}^2}} \right| \hat{\Theta} \sim N(0,1). \tag{A.2}$$

It follows immediately the third result of (i).

(ii) Based on (A.1) and (A.2) and the definition of $H_*(t)$, we have, for any $0 < s < 1$ and as $N \to \infty$,

$$\begin{aligned} &P\left\{H_*(\theta^{(m)}) \le s\right\} \\ &= P\left\{ P\left(\frac{\xi^* - \sum_{k=1}^K w_{k,(m)}\hat{\theta}_k / \sum_{k=1}^K w_{k,(m)}}{\sqrt{\sum_{k=1}^K w_{k,(m)}^2 s_k^2 / \{\sum_{k=1}^K w_{k,(m)}\}^2}} \le \left. \frac{\theta^{(m)} - \sum_{k=1}^K w_{k,(m)}\hat{\theta}_k / \sum_{k=1}^K w_{k,(m)}}{\sqrt{\sum_{k=1}^K w_{k,(m)}^2 s_k^2 / \{\sum_{k=1}^K w_{k,(m)}\}^2}} \right| \hat{\Theta} \right) \le s \right\} \\ &= P\left\{ \frac{\theta^{(m)} - \sum_{k=1}^K w_{k,(m)}\hat{\theta}_k / \sum_{k=1}^K w_{k,(m)}}{\sqrt{\sum_{k=1}^K w_{k,(m)}^2 s_k^2 / \{\sum_{k=1}^K w_{k,(m)}\}^2}} \le \Phi^{-1}(s) \right\} \to \Phi(\Phi^{-1}(s)) = s. \end{aligned}$$

Thus, $H_*(\theta^{(m)}) \to \mathrm{U}(0,1)$, as $N \to \infty$. The conclusion of (ii) follows.

A2. Proof of Lemma 3.1

Recall that the condition described in (4) is as follows:

$$\lim_{n\to\infty} w_{k,(m)} = \begin{cases} c_k & \text{if } k \in \Theta_T^{(m)}, \\ 0 & \text{if } k \notin \Theta_T^{(m)}, \end{cases} \quad \text{for } k = 1, 2, \ldots, K\ .$$

Without loss of generality, let $\theta_1 < \theta_2 < \ldots < \theta_K$. Also let $\hat{\theta}_k \sim N(\theta_k, \sigma_k^2/n_k)$ for each k and define $\hat{\theta}^{(k)}$ as $\hat{\theta}^{(1)} < \hat{\theta}^{(2)} < \ldots < \hat{\theta}^{(K)}$. Furthermore, suppose we are interested in θ_m. Recall $w_{m,(m)}^{[1]} = \mathbf{1}\{\hat{\theta}_m = \hat{\theta}^{(m)}\}$ is a binary random variable that equals 1 with probability

$P\{\hat{\theta}_m = \hat{\theta}^{(m)}\}$. Since $\Big\{\max_{k<m} \hat{\theta}_k < \hat{\theta}_m < \min_{k>m} \hat{\theta}_k\Big\} \subset \{\hat{\theta}_m = \hat{\theta}^{(m)}\}$, we have

$$
\begin{aligned}
P\{\hat{\theta}_m = \hat{\theta}^{(m)}\} &\geq \prod_{i<m} [P\{\hat{\theta}_i < \hat{\theta}_m\}] \prod_{j>m} [P\{\hat{\theta}_j > \hat{\theta}_m\}] \\
&= \int_{-\infty}^{\infty} \prod_{i<m} \left[\Phi(\frac{c-\theta_i}{\sigma_i/\sqrt{n_i}})\right] \prod_{j>m} \left[\Phi(\frac{\theta_j - c}{\sigma_j/\sqrt{n_j}})\right] \phi\left(\frac{c-\theta_m}{\sigma_m/\sqrt{n_m}}\right) \frac{dc}{\sigma_m/\sqrt{n_m}} \\
&\geq \int_{\theta_m-\epsilon_m}^{\theta_m+\epsilon_m} \prod_{k<m} \left[\Phi(\frac{c-\theta_i}{\sigma_i/\sqrt{n_i}})\right] \prod_{j>m} \left[\Phi(\frac{\theta_j - c}{\sigma_j/\sqrt{n_j}})\right] \phi\left(\frac{c-\theta_m}{\sigma_m/\sqrt{n_m}}\right) \frac{dc}{\sigma_m/\sqrt{n_m}} \\
&\geq \int_{\theta_m-\epsilon_m}^{\theta_m+\epsilon_m} \prod_{i\neq m} \left[\Phi\left(\frac{\sqrt{n_i}\epsilon_m}{2\sigma_i}\right)\right] \phi\left(\frac{c-\theta_m}{\sigma_m/\sqrt{n_m}}\right) \frac{dc}{\sigma_m/\sqrt{n_m}} \\
&= \int_{\theta_m-\epsilon_m}^{\theta_m+\epsilon_m} \Big\{1-o(1)\Big\} \phi\left(\frac{c-\theta_m}{\sigma_m/\sqrt{n_m}}\right) \frac{dc}{\sigma_m/\sqrt{n_m}} \\
&= \Big\{1-o(1)\Big\} \left\{\Phi\left(\frac{\sqrt{n_m}\epsilon_m}{\sigma_m}\right) - \Phi\left(-\frac{\sqrt{n_m}\epsilon_m}{\sigma_m}\right)\right\} \to 1
\end{aligned}
$$

for $\epsilon_m = \min\{(\theta_m - \theta_{m-1}), (\theta_{m+1} - \theta_m)\}/2$ as $N \to \infty$. Thus $w^{[1]}_{m,(m)}$ converges in probability to 1. Because we have that $w^{[1]}_{m,(m)} \to 1$ and $\sum_k w^{[1]}_{k,(m)} = 1$, $w^{[1]}_{k,(m)} \to 0$ $\forall k \neq m$, thus satisfying (4). Noting that $\hat{\theta}_k \sim N(\theta_k, \sigma_k^2/n_k)$ and, unconditionally, $\xi_k \sim N(\theta_k, 2\sigma_k^2/n_j)$, we can replace each σ_k^2 with $2\sigma_k^2$ in the proof above, and the result remains unchanged.

Recall that $w^{[3]}_{k,(m)} = \mathbf{1}\{-b_L \leq (\xi_k - \xi^{(m)}) \leq b_R\}$, where $(b_L, b_R) \propto \tau_N, \tau_N = O(N^{-\delta}), \delta \in (0, \frac{1}{2})$. For $i = m$, using the arguments above we have $P\{\xi_m = \xi^{(m)}\} \to 1$ and

$$
P\left\{\mathcal{K}\left(\frac{\xi_m - \xi^{(m)}}{\tau_N}\right) = \mathcal{K}\left(\frac{0}{\tau_N}\right) = 1\right\} \to 1 \quad \text{as} \quad N \to \infty.
$$

For $k \neq m$, $(\xi_k - \xi^{(m)})$ converges in probability to $D_k = \theta_k - \theta_m$. For $k < m$, $D_k/\tau_N \to -\infty$, and thus $\mathcal{K}\left(\frac{\xi_k - \xi^{(m)}}{\tau_N}\right) \to 0$. Similarly, for $k > m$, $D_k/\tau_N \to +\infty$, and thus $\mathcal{K}\left(\frac{\xi_k - \xi^{(m)}}{\tau_N}\right) \to 0$. Thus, we have verified (4).

A3. Proof of Lemma 3.2.

We only prove the general near tie case with $|\Theta^{(m)}_{\mathcal{N}}| = s \geq 1$; the exact tie case (i.e., $\Theta^{(m)}_{\mathcal{J}} = \Theta^{(m)}_{\mathcal{N}}$ case) is just a special case that can be proved by the same argument.

Without loss of generality, we assume that $\theta_1 \leq \cdots \leq \theta_{m_L-1} < \theta_{m_L} \leq \cdots \leq \theta_{m_U} < \theta_{m+1} \leq \cdots \leq \theta_K$ and $\Theta_{\mathcal{N}}^{(m)} = \{m_L, m_L+1, \cdots, m_U\}$, where $m_U - m_L = s \geq 1$ and $m_L \leq m \leq m_U$. As in the proof of Lemma 3.1, we write the ordered $\hat{\theta}_k$'s as $\hat{\theta}^{(1)} < \hat{\theta}^{(2)} < \cdots < \hat{\theta}^{(K)}$ and the ordered ξ_k's as $\xi^{(1)} < \xi^{(2)} < \cdots < \xi^{(K)}$. We also introduce the notation $\tilde{m}$ such that the m-th largest $\xi^{(m)}$ is from study $\tilde{m}$ with the underlying parameter $\theta_{\tilde{m}}$. Recall that $|\xi_k - \hat{\theta}_k| = O_p(N^{-1/2})$, $|\hat{\theta}_k - \theta_k| = O_p(N^{-1/2})$ and (thus) also $|\xi_k - \theta_k| = O_p(N^{-1/2})$, for all $k = 1, \ldots, K$.

We first prove that $\tilde{m} \in \Theta_{\mathcal{N}}^{(m)}$ with high probability for sufficiently large N. In particular, for each $k < m_L$ and $k' \geq m_L$, we have $\theta_{k'} - \theta_k \geq \theta_{m_L} - \theta_{m_L-1} \geq d_m$ and thus

$$\{\xi_{k'} - \xi_k\}/\tau_N = \{(\xi_{k'} - \theta_{k'}) - (\xi_k - \theta_k) + (\theta_{k'} - \theta_k)\}/\tau_N \geq \left\{1 - \frac{O_p(1)}{d_m N^{1/2}}\right\} \frac{d_m}{\tau_N} \to +\infty,$$

as $N \to \infty$. Therefore, we have $P(\xi_k < \xi_{k'}) \to 1$ as $N \to \infty$, for any $k < m_L$ and $k' \in \Theta_{\mathcal{N}}^{(m)}$. Similarly, on the upper part, we can prove that $P(\xi_k > \xi_{k'}) \to \infty$ as $N \to \infty$ for any $k > m_U$ and $k' \in \Theta_{\mathcal{N}}^{(m)}$. In general, we have

$$P\left(\max_{k<m_L} \xi_k < \min_{k \in \Theta_{\mathcal{N}}^{(m)}} \xi_k \leq \max_{k \in \Theta_{\mathcal{N}}^{(m)}} \xi_k < \min_{k>m_U} \xi_k\right) \to 1$$

as $N \to \infty$. Coupled with the fact that

$$\left\{\max_{k<m_L} \xi_k < \min_{k \in \Theta_{\mathcal{N}}^{(m)}} \xi_k \leq \max_{k \in \Theta_{\mathcal{N}}^{(m)}} \xi_k < \min_{k>m_U} \xi_k\right\} \subset \left\{\tilde{m} \in \Theta_{\mathcal{N}}^{(m)}\right\},$$

it is implied that

$$P\left(\tilde{m} \in \Theta_{\mathcal{N}}^{(m)}\right) \to 1 \quad \text{as} \quad N \to \infty.$$

Now noting the fact that $|\theta_k - \theta_{\tilde{m}}| \geq d_m$ for $\tilde{m} \in \Theta_{\mathcal{N}}^{(m)}$ and any $k \notin \Theta_{\mathcal{N}}^{(m)}$, we can show that for any $C > 0$

$$
\begin{aligned}
P\left(|\xi_k - \xi^{(m)}|/\tau_N \geq C\right) &= P\left(|\xi_k - \xi_{\tilde{m}}|/\tau_N \geq C\right) \\
&= P\left(|(\xi_k - \theta_k) - (\xi_{\tilde{m}} - \theta_{\tilde{m}}) + (\theta_k - \theta_{\tilde{m}})|/\tau_N \geq C\right) \\
&\geq P\left(\{|\theta_k - \theta_{\tilde{m}}| - |\xi_{\tilde{m}} - \theta_{\tilde{m}}| - |\xi_k - \theta_k|\}/\tau_N \geq C\right) \\
&\geq P\left(\{|\theta_k - \theta_{\tilde{m}}| - |\xi_{\tilde{m}} - \theta_{\tilde{m}}| - |\xi_k - \theta_k|\}/\tau_N \geq C, \tilde{m} \in \Theta_{\mathcal{N}}^{(m)}\right) \\
&\geq P\left(\left\{1 - \frac{O_p(1)}{d_m N^{1/2}}\right\}\frac{d_m}{\tau_N} \geq C, \tilde{m} \in \Theta_{\mathcal{N}}^{(m)}\right) \\
&\geq P\left(\left\{1 - \frac{O_p(1)}{d_m N^{1/2}}\right\}\frac{d_m}{\tau_N} \geq C\right) + P\left(\tilde{m} \in \Theta_{\mathcal{N}}^{(m)}\right) - 1 \\
&\to 1
\end{aligned}
$$

as $N \to \infty$. Because $b_L, b_R \propto \tau_N$, it follows that $P\left(w_{k,(m)}^{[3]} = \mathbf{1}\{-b_L \leq (\xi_k - \xi^{(m)}) \leq b_R\} = 0\right) \to 1$ for $k \notin \Theta_{\mathcal{N}}^{(m)}$ as $N \to \infty$.

Furthermore, for any $k \in \Theta_{\mathcal{N}}^{(m)}$, we have $|\theta_k - \theta_{\tilde{m}}| = O(N^{-1/2})$ by the near tie definition. It follows that for any $\epsilon > 0$

$$
\begin{aligned}
P\left(|\xi_k - \xi^{(m)}|/\tau_N \leq \epsilon\right) &= P\left(|\xi_k - \xi_{\tilde{m}}|/\tau_N \leq \epsilon\right) \\
&= P\left(|(\xi_k - \theta_k) - (\xi_{\tilde{m}} - \theta_{\tilde{m}}) + (\theta_k - \theta_{\tilde{m}})|/\tau_N \leq \epsilon\right) \\
&\geq P\left(\{|\theta_k - \theta_{\tilde{m}}| + |\xi_{\tilde{m}} - \theta_{\tilde{m}}| + |\xi_k - \theta_k|\}/\tau_N \leq \epsilon\right) \\
&\geq P\left(\{|\theta_k - \theta_{\tilde{m}}| + |\xi_{\tilde{m}} - \theta_{\tilde{m}}| + |\xi_k - \theta_k|\}/\tau_N \leq \epsilon, \tilde{m} \in \Theta_{\mathcal{N}}^{(m)}\right) \\
&\geq P\left(\frac{O_p(1)}{\tau_N N^{1/2}} \leq \epsilon, \tilde{m} \in \Theta_{\mathcal{N}}^{(m)}\right) \\
&\geq P\left(\frac{O_p(1)}{\tau_N N^{1/2}} \leq \epsilon\right) + P\left(\tilde{m} \in \Theta_{\mathcal{N}}^{(m)}\right) - 1 \\
&\to 1
\end{aligned}
$$

as $N \to \infty$. By noting that $b_L, b_R \propto \tau_N$, we have immediately $P\left(w_{k,(m)}^{[3]} = \mathbf{1}\{-b_L \leq (\xi_k - \xi^{(m)}) \leq b_R\} = 1\right) \to 1$ for any $k \in \Theta_{\mathcal{N}}^{(m)}$ as $N \to \infty$.

A4. Scale-invariant version of τ_N

Recall that $s^{(m)}$ is the standard error associated with $\hat{\theta}^{(m)}$, we may use the scale-invariant $(\sigma)(s^{(m)}/\sigma)^{2\delta}$ as τ_N for appropriate σ. For example, we may let $\sigma = \sqrt{(\sum_k s_k^2 n_k)/K} \propto 1$. Coupled with the fact that $s^{(m)} \propto N^{-1/2}$, this implies that $\tau_N \propto N^{-\delta}$.

A5. Implementing tuning of bandwidth parameters

Shrinkage of $\hat{\Theta}$ to obtain Θ^*:

For the purposes of selecting the appropriate bandwidth parameters (c_L, c_R), we propose a double bootstrap procedure, in which new sets of "observed data" are generated from a set of known parameters Θ^*. In order to prevent excess variability in Θ^* when $\hat{\Theta}$'s are likely to be unequal even when Θ's are equal, we shrink the observed estimates $\hat{\Theta}$, such that $\theta_k^* = \Delta \cdot \frac{\sum_{k=1}^{K} \hat{\theta}_k}{K} + (1-\Delta) \cdot \hat{\theta}_k$ for $k = 1, \ldots, K$ where $\Delta \in (0, 1]$. Intuitively, Δ should be large when the variation in $\hat{\Theta}$ is due primarily to within-study variation rather than true between-study variation, e.g. $\Delta \propto R = \frac{ss_{within}}{ss_{tot}} = \frac{\sum s_k^2}{\sum(\hat{\theta}_k - \bar{\hat{\theta}})^2} = \frac{\bar{s^2}}{var(\hat{\Theta})}$. It is easy to show that $R \to c > 0$ in probability under the fixed effects setting, while $R = o(1/N)$ otherwise. Because shrinkage is desirable under the fixed effects setting, we use $\Delta = max(1, R \cdot (\frac{\sigma}{\sqrt{\bar{s^2}}})^{0.1}) \approx max(1, R \cdot N^{0.05})$, where σ is as described in A4. Using this formulation, $R \to 1$ as $N \to \infty$ under the fixed-effects scenario and $R = o(N^{-0.95})$ otherwise.

Choosing (c_L, c_R) from a range of plausible values:

Sometimes, $\mathbb{L}(c_L, c_R)$ is insensitive over a certain region of (c_L, c_R). In this case, any (c_L, c_R) pairs such that $\mathbb{L}(c_L, c_R) < \gamma$ are considered admissible and the mean of all admissible pairs, denoted $(\bar{c}_L, \bar{c}_R)$, are chosen as our final bandwidth pair. Here, the threshold γ is the estimated $97.5th$ percentile of $\mathbb{L}_U = \frac{1}{B}\sum_{b=1}^{B}(\mathbb{U}_{(b)} - \frac{b}{B+1})^2$, where $\mathbb{U}_{(1)}, \ldots, \mathbb{U}_{(B)}$ are B ordered realizations from the standard U(0,1) distribution.

REFERENCES

Bagshaw, S. & Ghali, W. (2004). Acetylcysteine for prevention of contrast-induced nephropathy after intravascular angiography: a systematic review and meta-analysis. *BMC medicine* 2 38.

Brown, L. & Greenshtein, E. (2009). Nonparametric empirical bayes and compound decision approaches to estimation of high dimensional vector of normal means. *Annals of Statistics* 37 1685–1704.

Copas, J. (1969). Compound decisions and empirical bayes (with discussion). *Journal of Royal Statistical Society: Series B* 31 397–425.

Cox, D. R. (1958). Some problems connected with statistical inference. *The Annals of Mathematical Statististics* 29 357–372.

Cox, D. R. (2013). Discussion of "Confidence distribution, the frequentist distribution estimator of a parameter: a review". *International Statistical Review* 81 40–41.

Efron, B. (1993). Bayes and likelihood calculations from confidence intervals. *Biometrika* 80 3–26.

Fan, J. & Lv, J. (2011). Non-concave penalized likelihood with np-dimensionality. *IEEE transaction on Information Theory* 57 5467–5484.

Hall, P. & Miller, H. (2010). Bootstrap confidence intervals and hypothesis tests for extrema of parameters. *Biometrika* 97 881–892.

Hannig, J. & Xie, M. (2012). On Dempster-Shafer recombinations of confidence distributions. *Electrical Journal of Statistics* 6 1943–1966.

Jiang, W. & Zhang, C.-H. (2009). General maximum likelihood empirical bayes estimation of normal means. *Annals of Statistics* 37 1647–1684.

Schweder, T. & Hjort, N. (2002). Confidence and likelihood. *Scandinavian Journal of Statistics* 29 309–332.

Singh, K., Xie, M. & Strawderman, W. (2005). Combining information from independent sources through confidence distributions. *The Annals of Statistics* 33 159–183.

Small, C. G. (2010). *Expansions and Asymptotics for Statistics.* Chapman & Hall/CRC, New York.

Tian, L., Cai, T., Pfeffer, M., Piankov, N., Cremieux, P. & Wei, L. (2009). Exact and efficient inference procedure for meta-analysis and its application to the analysis of independent

2× 2 tables with all available data but without artificial continuity correction. *Biostatistics* 10 275–281.

van deer Vaart, A. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge UK.

Wandler, D. & Hannig, J. (2012). Generalized fiducial confidence intervals for extremes. *Extremes* 15 67–87.

Xie, M. & Singh, K. (2013). Confidence distribution, the frequentist distribution estimator of a parameter: a review (with discussions). *International Statistical Review* 81 3–39.

Xie, M., Singh, K. & Strawderman, W. E. (2011). Confidence distributions and a unified framework for meta-analysis. *Journal of the American Statistical Association* 106 320–333.

Xie, M., Singh, K. & Zhang, C.-H. (2009). Confidence intervals for population ranks in the presence of ties and near ties. *Journal of the American Statistical Association* 104 775–788.

Zhang, C.-H. (2003). Compound decision theory and empirical bayes methods. *Annals of Statistics* 31 379–390.