Harvard University
Harvard University Biostatistics Working Paper Series

Year 2012 *Paper* 154

Nonparametric Inference for Meta Analysis with Fixed Unknown, Study-specific Parameters

Brian Claggett* Minge Xie†
Lu Tian‡

*Harvard, bclagget@hsph.harvard.edu
†Rutgers University - New Brunswick/Piscataway, mxie@stat.rutgers.edu
‡Stanford University School of Medicine, lutian@stanford.edu

Nonparametric Inference for 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 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 unknown, fixed parameters, $\{\theta_1, \ldots, \theta_K\}$. Specifically, we draw inferences about, for example, the quantiles of this set of parameters using study-specific summary statistics. This type of problem is quite challenging (Hall and Miller, 2010). We utilize a novel resampling method via the confidence distributions of θ's 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 Postdoctoral Fellow, Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115 (E-mail: *bclagget@hsph.harvard.edu*). 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 Assistant Professor, Department of Health Research & Policy, Stanford University School of Medicine, Palo Alto, CA 94304 (E-mail: *lutian@stanford.edu*).

1. INTRODUCTION

Meta-analysis is a potentially powerful tool for combining information from multiple, independent studies for making inferences, 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 nonparametric framework without assuming that the underlying unknown parameters were realization of a random sample from a proper or degenerate distribution. Specifically, suppose that there are K independent studies whose *fixed*, unknown parameters are denoted by $\{\theta_1, \ldots, \theta_K\}$. The question is how to construct point and interval estimates, for example, for the $(100q)$th percentile, $\theta^{(q)}$, of this set of parameters via individual study-specific summary statistics.

Assume that, from the kth study, $k = 1, \ldots, K$, there is a $\sqrt{n_k}$- consistent estimator for θ_k, say $\hat{\theta}_k$, with a standard error estimate s_k, where n_k is the sample size for the kth study, $n = \sum_k n_k$, and the ratio n_k/n is stabilized away from 0 as $n \to \infty$. Moreover, assume that $(\hat{\theta}_k - \theta_k)/s_k$ is asymptotically normal. Furthermore, let $\Theta = \{\theta_1, \ldots, \theta_K\}$ and $\hat{\Theta} = \{\hat{\theta}_1, \ldots, \hat{\theta}_K\}$. 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 and Miller, 2010; Wandler and Hannig, 2012). In general, when several θ's are "clustered around" $\theta^{(q)}$, the inferential problem becomes non-trivial (Xie et al., 2009; Hall and Miller, 2010). The approach recommended by Hall and 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 bootstrap 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. This inference setup is associated with a well-known difficult problem (Hall and Miller, 2010; Wandler and Hannig, 2012). Hall and Miller (2010), who studied "the problem of constructing confidence intervals or hypothesis tests for extrema of parameters, for example of $\max\{\theta_1, \ldots, \theta_K\}$," 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 difficult part is the unknown 'ties' and 'near ties' cases. Here, a near tie case is defined as

$$|\theta_j - \theta^{(m)}| = O(N^{-1/2}),$$

which is interpreted as that, based on current sample size, a "near tie" parameter θ_j can not be distinguished from the target parameter $\theta^{(m)}$ (Xie et al., 2009; Hall and Miller, 2010).

In this paper, using the concept of confidence distributions (Xie and Singh, 2012), we propose a new and simple resampling method to construct confidence interval estimators for $\theta^{(q)}$, regardless of the presence or absence of '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 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 propose three reasonable weighting schemes, including the standard bootstrap estimator. In section 4, we discuss the properties of a set of weights which will guarantee appropriate asymptotic coverage, show how to construct weights that possess these properties, and discuss tuning approaches to improve finite-sample inference. In section 5, we present simulation results showing that our proposed weights provide 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 nonparametric inference framework for meta-analysis and also provides a solution for the well-established difficult problem of making inference for extrema of parameters.

2. CD-BASED INFERENCE

2.1 Introduction to Confidence Distribution

In frequentist inference, we often use a single sample statistic (point estimator) or an sample-dependent interval (confidence interval) to estimate a parameter of interest. A confidence distribution (CD) is quite similar, but uses a sample-dependent distribution function, rather than a single point (point estimator) or an interval (confidence

interval), to estimate the parameter of interest (Xie and Singh, 2012). A confidence distribution has also been loosely referred to as a sample-dependent distribution function that can represent confidence intervals of all levels for the parameter of interest (Cox, 1958; Efron, 1993). The concept has a long history, especially with its early interpretation associated with fiducial reasoning (Fisher, 1973; Cox, 2006). But recent developments have redefined a confidence distribution as a purely frequentist concept, without any fiducial reasoning. In his discussion of Xie and Singh (2012), Cox (2012) characterized the purely frequentist CD-based inference as an effective tool to provide "simple and interpretable summaries of what can reasonably be learned from data (and an assumed model)." Efron (2012) considered the recent CD development as "a grounding process" to solve "perhaps the most important unresolved problem in statistical inference" on "the use of Bayes theorem in the absence of prior information."

A simple example of a confidence distribution that has been broadly used in statistical practice is a bootstrap distribution. Efron (1998) explicitly stated that a bootstrap distribution is typically a "distribution estimator" and a "confidence distribution" function of the parameter that it targets. Singh et al. (2005, 2007) showed that a bootstrap distribution normally satisfies the definition of a confidence distribution or an asymptotic confidence distribution.

Another simple example, which is also used by Fisher (1930, 1973) to illustrate his fiducial function, is from the normal mean inference problem with sample $X_i \sim N(\mu, \sigma^2)$, $i = 1, \ldots, n$. The basic confidence distributions for μ are $\Phi(\sqrt{n}(\mu - \bar{X})/\sigma)$ when σ is known and $T_{n-1}(\sqrt{n}(\mu - \bar{x})/S)$ when σ is not known, and furthermore $\Phi(\sqrt{n}(\mu - \bar{X})/S)$ is an asymptotic confidence distribution when $n \to \infty$. Here, $\bar{X}$ and S^2 are the sample mean and variance, respectively, and T_{n-1} stands for the cumulative distribution function of the t-distribution with $n-1$ degrees of freedom. In other words, $N(\bar{X}, \sigma^2/n)$ is a "distribution estimator" of μ, when σ^2 is known. The distribution

functions $T_{n-1}(\sqrt{n}(\mu - \bar{x})/S)$ or $\Phi(\sqrt{n}(\mu - \bar{X})/S)$ can be used to estimate μ when σ^2 is not known. Similarly, in the context under consideration in this article, we can verify from Definition 1.1 of Singh et al. (2005) that

$$H_i(t) = \Phi\left(\frac{t - \hat{\theta}_i}{s_i}\right) \tag{1}$$

satisfies the requirements of being an asymptotic confidence distribution, thus we can use a distribution estimator $N(\hat{\theta}_i, s_i^2)$ to estimate θ_i, for $i = 1, 2, \ldots, K$.

A confidence distribution is a function of both the parameter and the random sample. For each given sample $\mathbf{X}_n$, a confidence distribution, say $H(\cdot) = H(\mathbf{X}_n, \cdot)$, is a cumulative distribution function on the parameter space; cf., Schweder and Hjort (2002); Singh et al. (2005). We can construct a random variable ξ defined on $\mathcal{X} \times \Xi$ such that, conditional on the sample data, ξ has the distribution H. Here, Ξ is the parameter space of the unknown parameter of interest θ and $\mathcal{X}$ is the sample space corresponding to data $\mathbf{X}_n = \{X_1, \ldots, X_n\}$. For example, let U be a $Unif[0, 1]$ random variable independent of $\mathbf{X}_n$, then $\xi = H^{-1}(U) \sim H(\cdot)$, given $\mathbf{X}_n$. We call this random variable ξ a *CD random variable* (Singh et al., 2007; Xie and Singh, 2012).

Definition 2.1. *We call $\xi = \xi_H$ a CD random variable associated with a confidence distribution H, if the conditional distribution of ξ given the data $\mathbf{X}_n$ is H.*

A CD random variable has a close association with a bootstrap estimator. In our example (1), a CD random variable ξ_i follows $\xi_i|\bar{x} \sim N(\hat{\theta}_i, s_i^2)$ and we have, asymptotically,

$$\frac{\xi_i - \hat{\theta}_i}{s_i}\,\Big|\,\hat{\theta}_i \;\sim\; \frac{\hat{\theta}_i - \theta_i}{s_i}\,\Big|\,\theta_i \qquad (\text{both} \sim N(0,1)).$$

This statement is exactly the same as the key justification for bootstrap, with ξ_i in place of the bootstrap sample mean $\hat{\theta}_i^*$. Thus, a CD random variable ξ can essentially

be viewed as a model-based bootstrap estimator of θ_i. Indeed, Xie and Singh (2012) 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 may inspire 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. The connection and the well-developed theory of bootstrap distributions can help us to understand inference procedures involving confidence distributions and develop new methodologies.

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 more liberally, which in turn allows us to develop more flexible statistical approaches and inference procedures.

3. PROPOSED METHODOLOGY

As illustrated in (1), from the ith study we have a confidence distribution (CD) function $H_i(t) = \Phi((t - \hat{\theta}_i)/s_i)$ that can be used to estimate θ_i, for $i = 1, \ldots, K$. Denote by ξ_i the CD random variable corresponding to $H_i(t) = \Phi((t - \hat{\theta}_i)/s_i)$, i.e.,

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

Given a particular realized set of $\{\xi_i, i = 1, \ldots, K\}$ from each of the K studies, we consider the construction of a weighted average of ξ_i's:

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

for the purposes of making inference on $\theta^{(m)}$. In particular, we can easily simulate $\{\xi_i, i = 1, \ldots, K\}$ 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 $\theta^{(m)}$.

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

Choice 1:

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

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

Choice 2:

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

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

Choice 3:

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

where $\mathcal{K}$ is a kernel function, and b_L, b_R represent the left-side and right-side kernel bandwidths. Without loss of generality, we henceforth assume a rectangular kernel, such that

$$\mathcal{K}\big(\xi_i - \xi^{(m)}, b_L, b_R\big) = \mathbf{1}\{-b_L \leq (\xi_i - \xi^{(m)}) \leq b_R\}.$$

Written this way, it is easy to see that $w_{i,(m)}^{[2]}$ represents a special case of $w_{i,(m)}^{[3]}$ in which

$b_L = b_R \equiv 0$.

Weights $w^{[1]}_{i,(m)}$ and $w^{[2]}_{i,(m)}$ both represent intuitively appealing ways of estimating and making inference on $\theta^{(m)}$. The use of $w^{[1]}_{i,(m)}$ 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 $w^{[2]}_{i,(m)}$ corresponds to the use of the distribution of the mth ordered ξ_i, and is therefore equivalent to the conventional bootstrap estimator of $\theta^{(m)}$, as discussed in Hall and 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. Our third option is flexible enough to appropriately handle a variety of scenarios while maintaining appropriate coverage levels, and in many cases, offering narrower confidence intervals than those obtained via the other weighting schemes. In the following section, we show that there is a very simple requirement for any given weighting scheme that allows for the use of ξ^* for asymptotically valid inference for $\theta^{(m)}$. Namely, $w_{i,(m)}$ must converge to a positive constant if θ_i belongs to the tie or near tie set of $\theta^{(m)}$, as defined below, and zero otherwise. We will show that this requirement is not satisfied by $w^{[1]}_{i,(m)}$ or $w^{[2]}_{i,(m)}$ when there are ties or near ties, but is satisfied by $w^{[3]}_{i,(m)}$ 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.

4. THEORETICAL RESULTS

First, let us formally define the tie and near tie sets. The same definition has also been utilized in Xie et al. (2009); Hall and Miller (2010). In particular, we denote by

$$\Theta^{(m)}_{\mathcal{T}} = \{j : \theta_j = \theta^{(m)}, j = 1, \ldots, K\}$$

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

$$\Theta_{\mathcal{N}}^{(m)} = \{j : |\theta_j - \theta^{(m)}| = O(N^{-1/2}), j = 1, \ldots, K\}$$

the "near tie set" of $\theta^{(m)}$. The interpretation of the "near tie" definition is that, based on current sample size n_i, a "near tie" parameter θ_i cannot be distinguished from the target parameter $\theta^{(m)}$. An equivalent expression is that, for any $j \in \Theta_{\mathcal{N}}^{(m)}$, $(\hat{\theta}_j - \hat{\theta}^{(m)}) - (\theta_j - \theta^{(m)}) \neq o_p(|\theta_j - \theta^{(m)}|)$, which means that the difference between θ_j and $\theta^{(m)}$ is not of greater order than the standard error of its estimator. 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: $\Theta_{\mathcal{T}}^{(m)} \subseteq \Theta_{\mathcal{N}}^{(m)}$. So, we also have the number of studies in the near tie set $|\Theta_{\mathcal{N}}^{(m)}| \geq |\Theta_{\mathcal{T}}^{(m)}| \geq 1$. 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 $\Theta_{\mathcal{T}}^{(m)}$.

4.1 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 carefully. In the theorem, Ξ is the parameter space of $\theta^{(m)}$. A proof of the theorem is provided in Appendix.

THEOREM 4.1. *Suppose that we can prove that a set of weights possesses the following property:*

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

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

(i)

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

where $\{s^{(m)}\}^2 = \sum_{i\in\Theta_{\mathcal{N}}^{(m)}} c_i^2 s_i^2 \big/ \big\{ \sum_{i\in\Theta_{\mathcal{N}}^{(m)}} c_i \big\}^2$. *Furthermore,*

$$\frac{\xi^* - \sum_{i=1}^{K} w_{i,(m)}\hat{\theta}_i / \sum_{i=1}^{K} w_{i,(m)}}{\sqrt{\sum_{i=1}^{K} w_{i,(m)}^2 s_i^2 / \{\sum_{i=1}^{K} w_{i,(m)}\}^2}} \Big| \hat{\Theta} \sim \frac{\sum_{i=1}^{K} w_{i,(m)}\hat{\theta}_i / \sum_{i=1}^{K} w_{i,(m)} - \theta^{(m)}}{\sqrt{\sum_{i=1}^{K} w_{i,(m)}^2 s_i^2 / \{\sum_{i=1}^{K} w_{i,(m)}\}^2}} \Big| \Theta, \qquad (5)$$

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

(ii) Denote by

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

When $t = \theta^{(m)}$, *we have* $H_*(\theta^{(m)}) \to Unif[0,1]$, *in distribution; Thus, by Definition 1.1 of Singh et al. (2005),* $H_*(\theta)$ *is an asymptotic CD for* $\theta^{(m)}$.

The function $H_*(t)$ is a cumulative distribution function on the parameter space Ξ and it also depends on the sample observations $\hat{\Theta}$. Based on the development on confidence distributions, the distribution estimator $H_*(\theta) = Pr(\xi^* \le \theta | \hat{\Theta})$ ensures asymptotically valid inference, including point estimation, confidence intervals, p-values, etc., for $\theta^{(m)}$; see, e.g., Singh et al. (2007); Xie and Singh (2012). Thus, in this case, we can rely on ξ^* to provide valid inference for $\theta^{(m)}$ asymptotically.

The remaining question is whether any of the three sets of weight choices in Section 3 satisfy the requirement (4) and, if they do, under which conditions. Since the asymptotic properties of each of the proposed weighted estimators depend on the true unknown values of Θ, we start 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.

Throughout the paper, we assume the following separation condition:

$$[\mathrm{C}_{sp}] \qquad d_m N^{1/2} \to \infty, \quad \text{where} \quad d_m = \min_{j \notin \Theta_{\mathcal{N}}^{(m)}} \left| \theta_j - \theta^{(m)} \right|$$

is the minimal distance between the θ_j's inside and outside the near tie set $\Theta_{\mathcal{N}}^{(m)}$. The separation condition $[\mathrm{C}_{sp}]$ allows that the separation d_m tends to zero but at a slower rate than $N^{-1/2}$. Condition $[\mathrm{C}_{sp}]$ is in fact weaker than the conventional assumption involving ties or no ties. Under the conventional setting, $d_m = \min_{\theta_j \neq \theta^{(m)}} |\theta_j - \theta^{(m)}| \geq c_o = \min_{i \neq j} |\theta_i - \theta_j|$ which is a typically positive constant bounded away from zero. Condition $[\mathrm{C}_{sp}]$ is also 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 we have no information regarding which ones are inside or outside the tie set.

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

Lemma 4.1 below states that, under the above no tie condition, all three choices of weights listed in Section 3 satisfy the condition in (4). A proof is given in the Appendix.

LEMMA 4.1 (ANY WEIGHT; NO TIE CASE). *Suppose that* $|\Theta_{\mathcal{N}}^{(m)}| = |\Theta_{\mathcal{T}}^{(m)}| = 1$ *and also Condition* $[C_{sp}]$ *holds. For* $s = 1, 2$, *we have*

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

Furthermore, if we use $w_{i,(m)}^{[3]}$ *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 $w^{[3]}_{i,(m)}$.

In conjunction with Theorem 4.1, 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 4.1, that the inference based on these three different choices of weights are asymptotically equivalent.

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

Example 4.1 (Counterexample for $w^{[1]}_{i,(m)}$ or $w^{[2]}_{i,(m)}$ in a simple tie case). Without loss of generality, consider a very simple example of a special case with $K = 2$ and $\theta_1 \equiv \theta_2$. For $m = 1$, $\Theta^{(m)}_{\mathcal{T}} = \Theta^{(m)}_{\mathcal{N}} = \{1, 2\}$ but $w^{[1]}_{1,(m)} = 1 - w^{[1]}_{2,(m)} = \mathbf{1}\{\hat{\theta}_1 = \min(\hat{\theta}_1, \hat{\theta}_2)\}$ is a binary random variable that equals 1 with probability $P\{\hat{\theta}_1 \le \hat{\theta}_2\} = 1 - P\{\hat{\theta}_2 \le \hat{\theta}_1\} = 0.5$. Thus, both $w^{[1]}_{1,(m)}$ and $w^{[1]}_{2,(m)}$ are (dependent) Bernoulli random variables, each with $p = 0.5$, therefore violating (4).

Similarly, for $m = 1$, the second choice of weights $w^{[2]}_{1,(m)} = 1 - w^{[2]}_{2,(m)} = \mathbf{1}\{\xi_1 = \min(\xi_1, \xi_2)\}$ is a binary random variable that equals 1 with probability $P\{\xi_1 \le \xi_2\} = E\big[P\{\xi_1 \le \xi_2 | \hat{\Theta}\}\big] = E\big[\Phi(\{\hat{\theta}_2 - \hat{\theta}_1\}/\{s_1^2 + s_2^2\}^{1/2})\big] = 0.5$. Again, both $w^{[2]}_{1,(m)}$ and $w^{[2]}_{2,(m)}$ are (dependent) Bernoulli random variables, each with $p = 0.5$, also violating (4).

In the case of more than two ties with either $|\Theta^{(m)}_{\mathcal{T}}| > 2$ or $|\Theta^{(m)}_{\mathcal{N}}| > 2$, the weights $w^{[1]}_{i,(m)}$ or $w^{[2]}_{i,(m)}$ for $i \in \Theta^{(m)}_{\mathcal{T}}$ or $\Theta^{(m)}_{\mathcal{N}}$ still converge to random quantities, rather than constants. The patterns are similar to, but more complicated than, those discussed in the case of $|\Theta^{(m)}_{\mathcal{T}}| = 2$ in Example 4.1. Clearly, neither $w^{[1]}_{i,(m)}$ nor $w^{[2]}_{i,(m)}$ satisfies the requirement (4) in this case, so we can no longer ensure that the results from Theorem 4.1 are valid. Our simulation results 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 $w_{i,(m)}^{[2]}$, was also reported by Hall and Miller (2010).

In contrast, if we use $w_{i,(m)}^{[3]}$ 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 requirement (4) is satisfied by $w_{i,(m)}^{[3]}$ 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 . We summarize the result in the following lemma, together with the result for a slightly modified $w_{i,(m)}^{[3]}$ choice:

$$\widetilde{w}_{i,(m)}^{[3]} = w_{i,(m)}^{[3]} \Big/ s_i.$$

A proof can be found in the Appendix.

LEMMA 4.2 (WEIGHT $w_{i,(m)}^{[3]}$; ANY CASE). *Suppose that Condition [C_{sp}] holds and we use $w_{i,(m)}^{[3]}$ with $b_L, b_R \propto \tau_N$, where $\tau_N/d_m \to 0$, and $\tau_N\sqrt{N} \to \infty$. For any $1 \le |\Theta_{\mathcal{J}}^{(m)}| \le |\Theta_{\mathcal{N}}^{(m)}| \le K$, we have*

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

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

This lemma, together with Theorem 4.1, provides a theoretical support to use the weighted sum of CD random variables ξ^* to make inference for $\theta^{(m)}$ in all cases, if either $w_{i,(m)}^{[3]}$ or $\widetilde{w}_{i,(m)}^{[3]}$ 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 $w_{i,(m)}^{[3]}$ is asymptotically equivalent to using the average of the $\hat{\theta}_i$ in the tie set (if we were

to know the true tie set). When s_i's or $\lambda_i = n_i/N$'s are heteroscedastic, the modified version $\widetilde{w}^{[3]}_{i,(m)}$ could be used to improve the efficiency and power of the inference. In fact, as in Theorem 2 of Xie et al. (2011), we can show that the inference based on $\tilde{w}^{[3]}_{i,(m)}$ is also asymptotically most efficient for $\theta^{(m)}$. 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 next in Section 4.2 on the tuning of the kernel widths.

4.2 Tuning the bandwidth parameters and a proposed algorithm

While we can guarantee that $w^{[3]}_{i,(m)}$ or $\tilde{w}^{[3]}_{i,(m)}$ 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 $|\Theta^{(m)}_T| = K$. 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 $w^{[2]}_{i,(m)}$, which we have shown to be asymptotically valid only when $|\Theta^{(m)}_T| = 1$. It seems reasonable that the tuning constants should be relatively large when ties are present and relatively small when no ties are present. Specifically, we refer to $\hat{\theta}_j : \{\hat{\theta}_j < \hat{\theta}^{(m)}, j \in \Theta_N\}$ as "left side ties" and $\hat{\theta}_j : \{\hat{\theta}_j > \hat{\theta}^{(m)}, j \in \Theta_N\}$ as "right side ties". We propose the usage of $(c_L,\ c_R)$ in order to ensure that the kernel smoother includes approximately the correct number of left and right side ties, respectively. This is a two-step process that involves estimating the potential number of left and right side ties, and then adjusting the kernel bandwidths to accommodate this estimated tie set.

We first attempt to detect the presence or absence of left side ties and/or right side ties by observing the behavior of the realized values of the CD random variables ξ. In general, we will simulate some large number, R, of samples of our CD random variables $\{\xi\}$. For $r = 1, 2, \ldots, R$, we may denote the r^{th} sampled value corresponding to $\hat{\theta}_i$ as $\xi_{i,r}$, the r^{th} collection of sampled values as $\{\xi_r\} = \{\xi_{i,r}, i = 1, \ldots, K\}$, and $\xi^{(m)}_r$ as the m^{th} smallest value among $\{\xi_{i,r}, i = 1, \ldots, K\}$. Here, without loss of generality, we assume $\hat{\theta}_1 \leq \hat{\theta}_2 \leq \ldots \leq \hat{\theta}_K$; otherwise, we can re-index the studies by the ordering of the $\hat{\theta}_i$ values. We define $\hat{\pi}_i = \frac{\sum_{r=1}^{R} \mathbf{1}\{\xi^{(m)}_r = \xi_{i,r}\}}{R}$. We may declare a particular study i to be a candidate for inclusion in the tie set if $\hat{\pi}_i$ exceeds some threshold γ near 0. The number of potential left side ties T^*_L is then given by $\sum_{i \leq m} I\{\hat{\pi}_i > \gamma\}$, and similarly, $T^*_R = \sum_{i \geq m} I\{\hat{\pi}_i > \gamma\}$ for the number of potential right side ties.

For a set of T studies with similar sample size, all sharing the same underlying study parameter θ, the maximum distance between any of the T values $\hat{\theta}$ is proportional to the range of T standard normal random variables, $R(T)$. We find that the $95th$ percentile of $R(T)$ can be well approximated by $R^*(T) = 2.58 + 0.79log(T)$, $T > 1$. Thus, we may use $c_L = R^*(T^*_L), c_R = R^*(T^*_R)$, with $R^*(1) = 0$.

However, it is important to recognize that T^*_L and T^*_R are only crude estimates

based on the data at hand, and may not reflect the true number of left and right side ties, particularly in small sample settings. As such, it is important to acknowledge the uncertainty in these estimates. Often, these estimates will serve as upper bounds for the true numbers of ties and we may perturb our constants (c_L, c_R) in order to reflect our uncertainty about the true number of ties, and thus the optimal smoothing bandwidths. To this end, we introduce a random tuning technique by setting $c_L = u \cdot R^*(T_L^*)$ and $c_R = u \cdot R^*(T_R^*)$, where u is a positive random number bounded between 0 and 1. In the simulation results presented, we generate $u \sim Unif(0, 1)$. We also implemented $u = \{0.001\}$ or $\{1\}$ with equal probability, and obtained similar results.

The introduction of the random tuning often increases the variability of the ξ^* in the proposed approach, thus effectively increasing the length of the reported confidence intervals. However, since u is bounded away from 0 and ∞, we can easily show that Lemma 4.2 still holds for this set of randomly tuned bandwidth choices. Indeed, with this set of bandwidth choices, there is no change in large sample properties and we observe very little change in numerical performance in large sample settings. However, under finite sample settings, this random tuning step accounts for additional variation and uncertainty due to lack of sample size and we observe a significant improvement in performance in our numerical studies.

Thus the algorithm for obtaining the CD used to estimate $\theta^{(m)}$ is as follows:

1. Calculate τ_N using observed $(\hat{\theta}, \hat{s}^2, n)$ data.
2. Generate $\{\xi_{i,r}, r = 1, \ldots, R\}$ for each study from $N(\hat{\theta}_i, \hat{s}_i^2)$; Here, without loss of generality, we assume $\hat{\theta}_1 \leq \hat{\theta}_2 \leq \ldots \leq \hat{\theta}_K$.
3. For each vector $\{\xi_r\}$, determine which study i is associated with $\xi_r^{(m)}$. We then use these counts to calculate $\hat{\pi}_i$ for each study, and functions thereof T_L^* and T_R^*.
4. For each r in $1, \ldots, R$

(a) Generate u_r from $Unif(0,1)$

(b) $\xi_r^* = \frac{\sum_i \xi_{i,r} \mathbf{1}\{-c_L \leq (\xi_{i,r} - \xi_r^{(m)})/\tau_N \leq c_R\}}{\sum_i \mathbf{1}\{-c_L \leq (\xi_{i,r} - \xi_r^{(m)})/\tau_N \leq c_R\}}$, with $c_L = u_r R^*(T_L^*)$ and $c_R = u_r R^*(T_R^*)$.

5. The CD for $\theta^{(m)}$ is approximated by the empirical distribution $\hat{H}_{\xi^*}(\theta)$, and a $(1-\alpha)100\%$ confidence interval can be estimated by $(\xi^*_{(R[\alpha/2])}, \xi^*_{(R[1-\alpha/2])})$.

5. SIMULATIONS

In order to demonstrate both small and large sample properties of our proposed estimator under different scenarios, we generate random data $X_{ij} \sim N(\theta_i, 1)$, with $\theta_i, i \in \{1, 2, \ldots, K\}, 1 \leq j \leq n_i$, taking different values according to the particular scenario:

1. Ties: $\theta_i = 0 \; \forall i$
2. Uniform: $\theta_i = \frac{2i}{K+1} - 1$
3. Normal: $\theta_i = \Phi^{-1}(\frac{i}{K+1})$

For each scenario, we consider $K = 7, 11$, or 21, and we let the sample size from each study $n_i = 40$, 400, or 4000. Using 500 simulated data sets for each setting, we show the coverage and median width of the nominal 95% and 80% confidence intervals.

We consider each of $w^{[1]}, w^{[2]}$, and $w^{[3]}$ as proposed in Section 2. The results are shown below. Because each set of $\{\Theta\}$ is symmetric, we need not show results for each ordered θ_i. Results are presented for the $5th$, $10th$, $25th$, and $50th$ percentiles. Note that for $K = 7$, the first ordered θ represents both the $5th$ and $10th$ percentile, and thus is shown only once. In particular, the coverage and median interval width for any $\theta^{(k)}$ will be identical to that for $\theta^{(K+1-k)}$.

For our proposed method using kernel smoothing, the results shown use the tuning procedure described in the previous section with R=1000 random samples drawn from each study's confidence distribution and using $\gamma = 0.05/K$ and $\delta = 2/5$. Simulation

results are shown below for $K = 7$ and 21. Simulation results corresponding to $K = 11$ show similar patterns and are available upon request.

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 which the coverage of Method 2 is far below the nominal level (e.g. the Ties setting, the Uniform setting with $n_i = 40$). This result matches the report of poor performance of the regular bootstrap approach in Hall and 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. Relative to the bootstrap estimator (Method 2), the intervals from our proposed method are 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 and 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 and Ghali (2004) resulted in a 95%

Table 1: Simulation results with K=7

			Method 1				Method 2				Method 3			
			95% interval		80% interval		95% interval		80% interval		95% interval		80% interval	
n_i	Quan	Dist	Cov	Wid	Cov	Wid	Cov	Wid	Cov	Wid	Cov	Wid	Cov	Wid
40	10th	1	0.820	0.613	0.422	0.404	0.068	0.454	0.012	0.295	0.920	0.564	0.692	0.375
	25th	1	0.998	0.608	0.846	0.400	0.512	0.355	0.190	0.232	0.962	0.404	0.826	0.263
	50th	1	1.000	0.617	0.990	0.403	0.992	0.316	0.908	0.207	0.986	0.311	0.896	0.195
	10th	2	0.962	0.609	0.810	0.401	0.960	0.544	0.814	0.357	0.954	0.681	0.808	0.453
	25th	2	0.976	0.605	0.856	0.398	0.978	0.488	0.858	0.320	0.966	0.613	0.848	0.401
	50th	2	0.974	0.611	0.862	0.400	0.988	0.479	0.874	0.316	0.974	0.612	0.848	0.404
	10th	3	0.952	0.607	0.774	0.400	0.952	0.586	0.788	0.387	0.950	0.664	0.750	0.438
	25th	3	0.954	0.606	0.804	0.399	0.960	0.546	0.806	0.361	0.950	0.670	0.792	0.447
	50th	3	0.964	0.613	0.832	0.400	0.972	0.513	0.844	0.338	0.956	0.637	0.816	0.424
400	10th	1	0.846	0.195	0.486	0.128	0.074	0.142	0.008	0.092	0.954	0.173	0.800	0.11
	25th	1	0.996	0.195	0.876	0.128	0.546	0.113	0.198	0.074	0.968	0.125	0.838	0.079
	50th	1	1.000	0.195	0.992	0.128	0.992	0.101	0.928	0.066	0.990	0.095	0.874	0.059
	10th	2	0.944	0.195	0.790	0.127	0.944	0.194	0.790	0.127	0.940	0.196	0.782	0.128
	25th	2	0.954	0.196	0.812	0.128	0.954	0.194	0.812	0.127	0.950	0.198	0.792	0.130
	50th	2	0.956	0.194	0.798	0.128	0.956	0.193	0.798	0.127	0.952	0.199	0.776	0.131
	10th	3	0.944	0.195	0.790	0.127	0.944	0.195	0.790	0.127	0.944	0.195	0.790	0.127
	25th	3	0.954	0.196	0.812	0.128	0.954	0.196	0.812	0.128	0.954	0.196	0.812	0.128
	50th	3	0.956	0.194	0.798	0.128	0.956	0.194	0.798	0.128	0.954	0.194	0.798	0.128
4000	10th	1	0.818	0.062	0.494	0.040	0.066	0.045	0.006	0.029	0.964	0.053	0.836	0.034
	25th	1	0.994	0.062	0.862	0.040	0.508	0.036	0.186	0.023	0.980	0.038	0.866	0.023
	50th	1	1.000	0.062	0.998	0.040	0.998	0.032	0.938	0.021	0.994	0.029	0.888	0.018
	10th	2	0.944	0.062	0.806	0.041	0.944	0.062	0.806	0.041	0.944	0.062	0.806	0.041
	25th	2	0.950	0.062	0.782	0.040	0.950	0.062	0.782	0.040	0.950	0.062	0.782	0.040
	50th	2	0.954	0.062	0.808	0.040	0.954	0.062	0.808	0.040	0.954	0.062	0.808	0.040
	10th	3	0.944	0.062	0.806	0.041	0.944	0.062	0.806	0.041	0.944	0.062	0.806	0.041
	25th	3	0.950	0.062	0.782	0.040	0.950	0.062	0.782	0.040	0.950	0.062	0.782	0.040
	50th	3	0.954	0.062	0.808	0.040	0.954	0.062	0.808	0.040	0.954	0.062	0.808	0.040

Table 2: Simulation results with K=21

n_i	Quan	Dist	Method 1 95% interval Cov	Method 1 95% interval Wid	Method 1 80% interval Cov	Method 1 80% interval Wid	Method 2 95% interval Cov	Method 2 95% interval Wid	Method 2 80% interval Cov	Method 2 80% interval Wid	Method 3 95% interval Cov	Method 3 95% interval Wid	Method 3 80% interval Cov	Method 3 80% interval Wid
40	5th	1	0.878	0.611	0.336	0.402	0.000	0.300	0.000	0.194	0.908	0.449	0.694	0.321
	10th	1	0.956	0.613	0.636	0.401	0.000	0.256	0.000	0.168	0.928	0.378	0.742	0.267
	25th	1	1.000	0.609	0.968	0.398	0.144	0.207	0.018	0.136	0.946	0.259	0.802	0.171
	50th	1	1.000	0.616	1.000	0.402	0.986	0.189	0.908	0.124	0.984	0.179	0.868	0.111
	5th	2	0.988	0.608	0.950	0.400	0.978	0.390	0.838	0.255	0.992	0.602	0.882	0.419
	10th	2	0.998	0.608	0.976	0.400	0.996	0.357	0.904	0.234	0.996	0.543	0.878	0.372
	25th	2	1.000	0.608	0.968	0.398	0.996	0.329	0.928	0.217	0.996	0.456	0.906	0.303
	50th	2	1.000	0.612	0.972	0.402	0.996	0.328	0.928	0.216	0.992	0.432	0.914	0.282
	5th	3	0.964	0.606	0.832	0.398	0.970	0.498	0.828	0.328	0.968	0.692	0.854	0.463
	10th	3	0.988	0.604	0.886	0.397	0.984	0.456	0.878	0.300	0.988	0.642	0.860	0.424
	25th	3	0.996	0.612	0.940	0.401	0.992	0.394	0.906	0.259	0.992	0.539	0.892	0.358
	50th	3	0.996	0.608	0.958	0.401	0.992	0.365	0.930	0.237	0.994	0.479	0.914	0.313
400	5th	1	0.898	0.196	0.334	0.128	0.000	0.094	0.000	0.061	0.950	0.140	0.804	0.098
	10th	1	0.992	0.195	0.684	0.127	0.000	0.081	0.000	0.053	0.950	0.117	0.816	0.081
	25th	1	1.000	0.195	0.988	0.128	0.120	0.066	0.008	0.043	0.966	0.081	0.856	0.051
	50th	1	1.000	0.195	1.000	0.128	0.996	0.060	0.928	0.039	0.986	0.054	0.892	0.034
	5th	2	0.972	0.195	0.860	0.128	0.970	0.162	0.878	0.107	0.976	0.213	0.840	0.142
	10th	2	0.972	0.195	0.866	0.128	0.978	0.159	0.866	0.105	0.964	0.208	0.850	0.136
	25th	2	0.978	0.195	0.832	0.128	0.984	0.160	0.844	0.106	0.970	0.223	0.836	0.146
	50th	2	0.988	0.194	0.872	0.128	0.992	0.159	0.888	0.105	0.980	0.223	0.874	0.146
	5th	3	0.952	0.195	0.810	0.128	0.952	0.194	0.810	0.127	0.950	0.200	0.794	0.130
	10th	3	0.958	0.194	0.822	0.128	0.958	0.191	0.822	0.126	0.946	0.231	0.784	0.157
	25th	3	0.962	0.195	0.796	0.128	0.964	0.179	0.810	0.119	0.942	0.240	0.778	0.159
	50th	3	0.972	0.194	0.852	0.128	0.980	0.168	0.866	0.112	0.956	0.231	0.832	0.153
4000	5th	1	0.914	0.062	0.374	0.040	0.000	0.030	0.000	0.019	0.956	0.043	0.846	0.029
	10th	1	0.986	0.062	0.654	0.040	0.000	0.026	0.000	0.017	0.958	0.036	0.852	0.024
	25th	1	1.000	0.062	0.982	0.040	0.138	0.021	0.020	0.014	0.966	0.025	0.868	0.015
	50th	1	1.000	0.062	1.000	0.040	0.998	0.019	0.926	0.013	0.982	0.016	0.876	0.010
	5th	2	0.974	0.062	0.816	0.04	0.974	0.062	0.816	0.04	0.962	0.062	0.788	0.041
	10th	2	0.952	0.062	0.788	0.041	0.952	0.062	0.788	0.04	0.938	0.062	0.764	0.041
	25th	2	0.95	0.062	0.804	0.04	0.95	0.061	0.804	0.04	0.938	0.062	0.774	0.041
	50th	2	0.958	0.062	0.814	0.04	0.958	0.061	0.814	0.04	0.952	0.062	0.794	0.041
	5th	3	0.974	0.062	0.816	0.04	0.974	0.062	0.816	0.04	0.974	0.062	0.816	0.04
	10th	3	0.952	0.062	0.788	0.041	0.952	0.062	0.788	0.041	0.952	0.062	0.788	0.041
	25th	3	0.95	0.062	0.804	0.04	0.95	0.062	0.804	0.04	0.95	0.062	0.804	0.04
	50th	3	0.958	0.062	0.814	0.04	0.958	0.062	0.814	0.04	0.956	0.062	0.812	0.04

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)

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 and Ghali (2004), assuming that 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. 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 log-odds-ratios 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 $w^{[3]}$ (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 (0.25, 1.01) and (0.31, 1.11), 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 (0.28, 1.06) 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 confidence distribution. 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)	(0.01, 0.60)
2	(0.02, 0.51)	(0.03, 0.20)	(0.04, 0.64)
3	(0.01, 0.94)	(0.05, 0.28)	(0.07, 0.68)
4	(0.01, 4.64)	(0.07, 0.40)	(0.11, 0.70)
5	(0.04, 1.04)	(0.12, 0.54)	(0.14, 0.75)
6	(0.09, 0.94)	(0.17, 0.70)	(0.19, 0.86)
7	(0.19, 1.67)	(0.25, 0.91)	(0.25, 1.01)
8	(0.10, 3.85)	(0.33, 1.16)	(0.31, 1.11)
9	(0.27, 4.93)	(0.45, 1.46)	(0.32, 1.33)
10	(0.39, 3.94)	(0.56, 1.79)	(0.35, 1.65)
11	(0.45, 3.49)	(0.70, 2.27)	(0.38, 1.94)
12	(0.26, 6.06)	(0.87, 2.99)	(0.34, 2.27)
13	(0.28, 6.14)	(1.09, 4.38)	(0.35, 3.05)
14	(0.44, 4.26)	(1.44, 9.56)	(0.42, 6.17)

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 CDF 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 $F(\Theta)$, 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 exact 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 1000 times from the exact 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. We examined the $25th$, $50th$, and $75th$ percentiles of the study-specific parameters, and found that both the 95% and 80% 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 $(-0.21, 0.49)\%$ for the "median" treatment effect, with intervals of $(-0.46, 0.49)\%$ and $(-0.13, 0.91)\%$ for the $25th$ and $75th$ percentiles, respectively.

7. DISCUSSION

In this paper, we introduce a unified framework which simultaneously addresses two important problems. By introducing a procedure for making inference on *any* ordered

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

		95% interval		80% interval	
Quantile	**OS**	**Cov**	**Wid**	**Cov**	**Wid**
Method 1					
25th	**8**	1.000	0.049	1.000	0.029
50th	**15**	1.000	0.040	1.000	0.024
75th	**22**	1.000	0.048	0.998	0.027
Method 2					
25th	**8**	0.888	0.019	0.532	0.012
50th	**15**	0.986	0.012	0.886	0.007
75th	**22**	0.734	0.017	0.316	0.011
Method 3					
25th	**8**	0.994	0.022	0.830	0.013
50th	**15**	0.994	0.014	0.930	0.009
75th	**22**	0.998	0.021	0.904	0.012

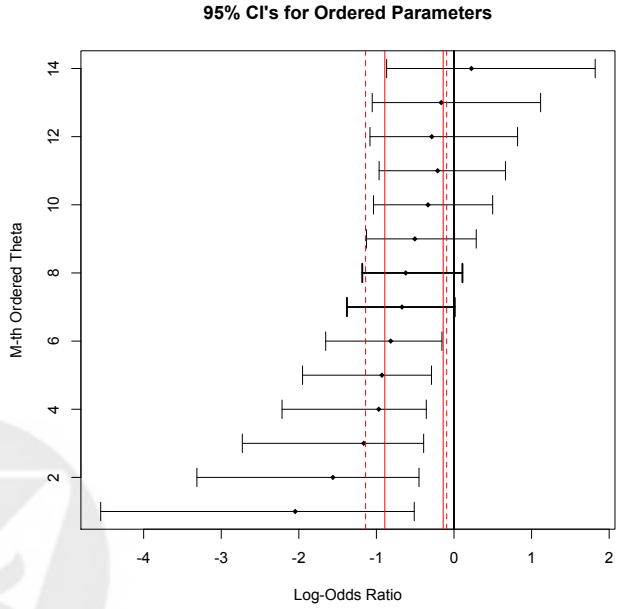

Figure 1: Confidence distribution estimates of treatment effects from 14 studies of acetylcysteine on nephropathy: Vertical solid (dashed) lines represent 95% CI from fixed-effects (random-effects) meta-analysis

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 nonparametric, model-free form of meta-analysis. While the resulting confidence interval from such a procedure will likely 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. 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.

APPENDIX

A1. Proof of Theorem 4.1.

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

Note that, $\hat{\theta}_i \sim (\theta_i, s_i^2)$, for any i, it follows that

$$\frac{\sum_{i\in\Theta_{\mathcal{N}}} c_i\hat{\theta}_i / \sum_{i\in\Theta_{\mathcal{N}}} c_i - \sum_{i\in\Theta_{\mathcal{N}}} c_i\theta_i / \sum_{i\in\Theta_{\mathcal{N}}} c_i}{\sqrt{\sum_{i\in\Theta_{\mathcal{N}}} c_i^2 s_i^2 / \{\sum_{i\in\Theta_{\mathcal{N}}} c_i\}^2}} \sim N(0,1).$$

Again, from (4) and the fact that $|\theta_i - \theta^{(m)}| = O(N^{-1/2})$ for any $\theta_i \in \Theta_{\mathcal{N}}$, we have $\sum_{i=1}^K w_{i,(m)}\hat{\theta}_i = \sum_{i\in\Theta_{\mathcal{N}}} c_i\hat{\theta}_i + o_p(1)$, $\sum_{i=1}^K w_{i,(m)}^2 s_i^2 = \sum_{i\in\Theta_{\mathcal{N}}} c_i^2 s_i^2 + o_p(1)$, $\sum_{i=1}^K w_{i,(m)} =$

$\sum_{i\in\Theta_N} c_i + o_p(1)$ and $\sum_{i\in\Theta_N} c_i\theta_i = \{\sum_{i\in\Theta_N} c_i\}\,\theta^{(m)} + O(N^{-1/2})$. Thus, we have

$$\frac{\sum_{i=1}^K w_{i,(m)}\hat{\theta}_i/\sum_{i=1}^K w_{i,(m)} - \theta^{(m)}}{\sqrt{\sum_{i=1}^K w_{i,(m)}^2 s_i^2/\{\sum_{i=1}^K w_{i,(m)}\}^2}} \to N(0,1), \quad \text{as } N\to\infty \tag{A.1}$$

On the other hand, since $\xi^* = \sum_{i=1}^K w_{i,(m)}\xi_i/\sum_{i=1}^K w_{i,(m)}$ and ξ_i are CD random variables from $N(\hat{\theta}_i, s_i^2)$, we have

$$\frac{\xi^* - \sum_{i=1}^K w_{i,(m)}\hat{\theta}_i/\sum_{i=1}^K w_{i,(m)}}{\sqrt{\sum_{i=1}^K w_{i,(m)}^2 s_i^2/\{\sum_{i=1}^K w_{i,(m)}\}^2}}\Big|\hat{\Theta} \sim N(0,1). \tag{A.2}$$

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 $N\to\infty$,

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

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

A2. Proof of Lemma 4.1

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

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

Without loss of generality, let $\theta_1 < \theta_2 < \ldots < \theta_K$. Also let $\hat{\theta}_j \sim N(\theta_j, \sigma_j^2/n_j)$ for each j and define $\hat{\theta}^{(j)} : \hat{\theta}^{(1)} \le \hat{\theta}^{(2)} \le \ldots \le \hat{\theta}^{(K)}$. Furthermore, suppose we are interested in θ_m.

Recall $w^{[1]}_{m,(m)} = \mathbf{1}\{\hat{\theta}_m = \hat{\theta}^{(m)}\}$ is a binary random variable that equals 1 with probability $P\{\hat{\theta}_i = \hat{\theta}^{(m)}\}$.

$$
\begin{aligned}
P\{\hat{\theta}_m = \hat{\theta}^{(m)}\} &< \prod_{i<m}[P\{\hat{\theta}_i < \hat{\theta}_m\}] \prod_{j>m}[P\{\hat{\theta}_j > \hat{\theta}_m\}] \\
&= \int \prod_{i<m}[P\{\hat{\theta}_i < c\}] P\{\hat{\theta}_m = c\} \prod_{j>m}[P\{\hat{\theta}_j > c\}]\,\mathrm{d}c \\
&= \int \prod_{i<m}[\Phi(\frac{c-\theta_i}{\sigma_i/\sqrt{n_i}})]\phi(\frac{c-\theta_m}{\sigma_m/\sqrt{n_m}}) \prod_{j>m}[\Phi(\frac{\theta_j - c}{\sigma_j/\sqrt{n_j}})]\,\mathrm{d}c \\
&< \int_{\theta_{m-1}+\epsilon}^{\theta_{m+1}-\epsilon} \prod_{i<m}[\Phi(\frac{c-\theta_i}{\sigma_i/\sqrt{n_i}})]\phi(\frac{c-\theta_m}{\sigma_m/\sqrt{n_m}}) \prod_{j>m}[\Phi(\frac{\theta_j - c}{\sigma_j/\sqrt{n_j}})]\,\mathrm{d}c \\
&\to \int_{\theta_{m-1}+\epsilon}^{\theta_{m+1}-\epsilon} \phi(\frac{c-\theta_m}{\sigma_m/\sqrt{n_m}})\,\mathrm{d}c \to 1
\end{aligned}
$$

Thus $w^{[1]}_{m,(m)}$ converges in probability to 1. Because we have that $w^{[1]}_{m,(m)} \to 1$ and $\sum_i w^{[1]}_{i,(m)} = 1$, then $w^{[1]}_{i,(m)} \to 0 \ \forall i \neq m$, thus satisfying (4). Noting that $\hat{\theta}_j \sim N(\theta_j, \sigma_j^2/n_j)$ and, unconditionally, $\xi_j \sim N(\theta_j, 2\sigma_j^2/n_j)$, we can replace each σ_j^2 with $2\sigma_j^2$ in the proof above, and the result remains unchanged.

Recall that $w^{[3]}_{i,(m)} = \mathbf{1}\{-b_L \leq (\xi_i - \xi^{(m)}) \leq b_R\}$, where $(b_L, b_R) \propto \tau_N, \tau_N = O(N^{-\delta}), \delta \in (0, \frac{1}{2})$. For $i = m$, we use the argument above that $P\{\xi_m = \xi^{(m)}\} \to 1$, and so $\mathcal{K}\left(\frac{\xi_i - \xi^{(m)}}{\tau_N}\right) \to \mathcal{K}\left(\frac{0}{\tau_N}\right) = 1$. For $i \neq m$, $(\xi_i - \xi^{(m)})$ converges in probability to $D_i = \theta_i - \theta_m$. For $i < m$, $D_i/\tau_N \to -\infty$, and thus $\mathcal{K}\left(\frac{\xi_i - \xi^{(m)}}{\tau_N}\right) \to 0$. Similarly, for $i > m$, $D_i/\tau_N \to +\infty$, and thus $\mathcal{K}\left(\frac{\xi_i - \xi^{(m)}}{\tau_N}\right) \to 0$. Thus, we have satisfied (4).

A3. Proof of Lemma 4.2.

Recall that $w^{[3]}_{i,(m)} = \mathbf{1}\{-b_L \leq (\xi_i - \xi^{(m)}) \leq b_R\}$, where $(b_L, b_R) \propto \tau_N$, and $\tau_N = O(N^{-\delta})$, with $0 < \delta < 1/2$. Let us denote $c_L = b_L/\tau_N, c_R = b_R/\tau_N$, so that $c_L, c_R = O(1)$.

Now $w^{[3]}_{i,(m)} = \mathbf{1}\{-b_L \leq (\xi_i - \xi^{(m)}) \leq b_R\} = \mathbf{1}\{-c_L \leq \frac{(\xi_i - \xi^{(m)})}{\tau_N} \leq c_R\}$.

Note that in general, $\xi_i = \hat{\theta}_i + \tilde{\epsilon}_n$, and $\hat{\theta}_i = \theta_i + \epsilon_n$, where both $\tilde{\epsilon}_n$ and $\epsilon_n = O(N^{-1/2})$. Substituting, we have that $(\xi_i - \xi^{(m)}) = \theta_i - \theta^{(m)} + \tilde{\epsilon}_{in} - \tilde{\epsilon}_n^{(m)} + \epsilon_{in} - \epsilon_n^{(m)} = \theta_i - \theta^{(m)} + O(N^{-1/2})$.

Thus, when $i \in \Theta_{\mathcal{N}}^{(m)}, \theta_i = \theta^{(m)} + o(N^{-1/2})$, then $(\xi_i - \xi^{(m)}) = O(N^{-1/2}), \frac{(\xi_i - \xi^{(m)})}{\tau_N} = O(N^{\delta - 1/2}) = o(1) \Rightarrow \frac{(\xi_i - \xi^{(m)})}{\tau_N} \to 0$ as $N \to \infty$, and so $w_{i,(m)}^{[3]} \to 1$ as $N \to \infty$.

Recalling that $d_m = \min_{j \notin \Theta_T^{(m)}} |\theta_j - \theta^{(m)}|$ is $O(N^{-\delta^*})$, for any $\delta^* \in [0, 1/2)$, we require that the convergence rate for τ_N, δ must be restricted to $(\delta^*, 1/2)$.

When $i \notin \Theta_{\mathcal{N}}^{(m)}, \theta_i \neq \theta^{(m)}$, then $(\xi_i - \xi^{(m)}) = O(N^{-\delta^*}), \frac{(\xi_i - \xi^{(m)})}{\tau_N} = O(N^{\delta - \delta^*}) \Rightarrow \frac{(\xi_i - \xi^{(m)})}{\tau_N} \to \infty$ as $N \to \infty$, and so $w_{i,(m)}^{[3]} \to 0$ as $N \to \infty$.

A4. Scale-invariant version of τ_N

We may instead use $\tau_N = (\sigma)(s^{(m)}/\sigma)^{2\delta}$, where $s^{(m)}$ is the standard error associated with $\hat{\theta}^{(m)}$ and σ is reasonable maximum value for the tuning parameter, such as $\sigma = \sqrt{\frac{\sum_i s_i^2 n_i}{K}}$. This particular formulation of σ ensures that $s^{(m)}/\sigma \approx n_{(m)}^{-1/2}$, and that $(s^{(m)}/\sigma)^{2\delta} > (s^{(m)}/\sigma)$. Note that $\sigma = O(1), s^{(m)} = O(N^{-1/2})$, and so $\tau_N = O(-\delta)$.

REFERENCES

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

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

Cox, D. (2006). Some problems with statistical inference. *London: Cambridge University Press* .

Cox, D. R. (in press, 2012). Discussions on the article "confidence distribution, the frequentist distribution estimator of a parameter". *International Statistical Review* .

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

Efron, B. (1998). R.A.Fisher in the 21st century. *Statistical Science* **13,** 95–122.

Efron, B. (in press, 2012). Discussions on the article "confidence distribution, the frequentist distribution estimator of a parameter". *International Statistical Review* .

Fisher, R. (1930). Inverse probability. *Proc. Cambridge Pilos. Soc.* **26,** 528–535.

Fisher, R. (1973). Statistical methods and scientific inference. *New York: Hafner Press* .

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

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

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

Singh, K., Xie, M., and Strawderman, W. (2007). Confidence distribution (cd): Distribution estimator of a parameter. In R. Liu, et al., editor, *Complex Datasets and Inverse Problems*, IMS Lecture Notes-Monograph Series No. 54, pages 132–150. IMS, Beachwood, Ohio, USA.

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

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

Xie, M. and Singh, K. (in press, 2012). Confidence distribution, the frequentist distribution estimator of a parameter — a review. *International Statistical Review* .

Xie, M., Singh, K., and 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., and 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.