Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: http://www.tandfonline.com/loi/uasa20

# Confidence Distributions and a Unifying Framework for Meta-Analysis

Minge Xie, Kesar Singh & William E. Strawderman

**To cite this article:** Minge Xie, Kesar Singh & William E. Strawderman (2011) Confidence Distributions and a Unifying Framework for Meta-Analysis, Journal of the American Statistical Association, 106:493, 320-333, DOI: 10.1198/jasa.2011.tm09803

**To link to this article:** http://dx.doi.org/10.1198/jasa.2011.tm09803

Published online: 01 Jan 2012.

# Confidence Distributions and a Unifying Framework for Meta-Analysis

Minge XIE, Kesar SINGH, and William E. STRAWDERMAN

This article develops a unifying framework, as well as robust meta-analysis approaches, for combining studies from independent sources. The device used in this combination is a confidence distribution (CD), which uses a distribution function, instead of a point (point estimator) or an interval (confidence interval), to estimate a parameter of interest. A CD function contains a wealth of information for inferences, and it is a useful device for combining studies from different sources. The proposed combining framework not only unifies most existing meta-analysis approaches, but also leads to development of new approaches. We illustrate in this article that this combining framework can include both the classical methods of combining $p$-values and modern model-based meta-analysis approaches. We also develop, under the unifying framework, two new robust meta-analysis approaches, with supporting asymptotic theory. In one approach each study size goes to infinity, and in the other approach the number of studies goes to infinity. Our theoretical development suggests that both these robust meta-analysis approaches have high breakdown points and are highly efficient for normal models. The new methodologies are applied to study-level data from publications on prophylactic use of lidocaine in heart attacks and a treatment of stomach ulcers. The robust methods performed well when data are contaminated and have realistic sample sizes and number of studies.

KEY WORDS: Combination of $p$-values; Fixed-effects model; Random-effects model; Robust methods.



## 1. INTRODUCTION

In the modern era with explosive growth of information, it is important to process information in an efficient and meaningful manner. Meta-analysis is such a statistical methodology that combines results from separate studies. The topic of meta-analysis has an enormous literature. For instance, the review of recent developments by Sutton and Higgins (2008) on model-based meta-analysis alone listed 281 references. Some recent books include, for example, Hedges and Olkin (1985), Stangl and Berry (2000), Whitehead (2002), Schulze (2004), Preiss et al. (2006), and many more. Indeed, collecting together overall information from different studies is a critical component for decision making. Combined results from multiple studies summarize overall associations, and inferences from the combined results are typically more reliable than inferences from any single study. The study of formal and meaningful ways of combining studies from independent sources is important both theoretically and practically. This article develops a unifying framework for combining studies from independent sources and, based on this framework, develops robust meta-analysis approaches that can effectively mitigate the impact of outlying studies.

The device used in our proposed combination approach is a confidence distribution (CD), a concept loosely referring to a distribution function that can represent confidence intervals of all levels for a parameter of interest. The CD concept has a long history (see, e.g., Fisher 1973 and Efron 1993), but recent developments have redefined the CD as a *purely frequentist concept* and focused on providing inference tools for problems in modern applied statistics. Generally speaking, a CD approach uses a distribution function, instead of a point or an interval, to estimate a parameter of interest. More specifically, let $\Theta$ be the parameter space of a parameter of interest $\theta$ and $\mathcal{X}$ be the sample space of the sample observations $\mathbf{X} = \{X_1, X_2, \ldots, X_n\}$. A CD function $H(\cdot) = H(\mathbf{X}, \cdot)$ is a mapping from $\mathcal{X} \times \Theta \to [0, 1]$ where, for each given sample $\mathbf{X} \in \mathcal{X}$, $H(\cdot)$ is a sample-dependent continuous cumulative distribution function on $\Theta$. Also, we require that, when $\theta = \theta_0$ the true parameter value, $H(\theta_0) \equiv H(\mathbf{X}, \theta_0)$, as a function of the sample $\mathbf{X}$, follows the uniform distribution U[0, 1]. The U[0, 1] requirement guarantees that inferences (such as confidence intervals, point estimators, $p$-values, etc.) derived from the CD function have desired properties for making inference on $\theta_0$. The function $H(\cdot)$ is an asymptotic CD (aCD), if this U[0, 1] requirement is true only asymptotically. See Definition A.1 in Appendix A.1, which was formulated in Schweder and Hjort (2002) and Singh, Xie, and Strawderman (2005). This new CD definition is consistent with the classical CD notion which is compiled from confidence intervals of varying confidence levels (cf., Singh, Xie, and Strawderman 2005). A CD function contains a wealth of information for inferences; much more than a point estimator or a confidence interval. Schweder and Hjort (2003) suggested that a CD is a "frequentist analogue of a Bayesian posterior." However, the notion of CD, especially in its asymptotic form, is much broader. Recent research has shown that the new CD concept encompasses and unifies a wide range of examples, from regular parametric cases (including most examples in the classical development of Fisher's fiducial distributions) to bootstrap distributions, significance ($p$-value) functions (Fraser 1991), normalized likelihood functions, and, in some cases, Bayesian priors and Bayesian posteriors. There is renewed interest in CDs (e.g., Efron 1998; Schweder and Hjort 2002; Schweder 2003; Lawless and Fredette 2005; Parzen 2005; Singh, Xie, and Strawderman 2005, 2007; Xie et al. 2009; Tian et al. 2010; among others). A brief review and highlights of recent developments on CDs are provided in Appendix A.2.

Minge Xie (E-mail: *mxie@stat.rutgers.edu*), Kesar Singh (E-mail: *kesar@stat.rutgers.edu*), and William E. Strawderman (E-mail: *straw@stat.rutgers.edu*) are Professors of Statistics, Department of Statistics and Biostatistics, Rutgers University, Piscataway, NJ 08854. This research is partly supported by research grants from NSF (DMS0915139, SES0851521) and NSA (H98230-08-1-0104). The authors are very grateful to Brad Efron, Ingram Olkin, Regina Liu, Jan Hannig, and David Hoaglin, as well as the editors and two anonymous reviewers, for their constructive suggestions and discussions.


Journal of the American Statistical Association
March 2011, Vol. 106, No. 493, Theory and Methods
DOI: 10.1198/jasa.2011.tm09803

In this article, we extend the general recipe for combining CDs proposed by Singh, Xie, and Strawderman (2005) to meta-analysis settings and develop a general framework to perform meta-analysis from separate studies. With the flexibility of the general framework and the breadth of the CD concept, the proposed CD-combining procedure represents a large class of combining methods for meta-analysis. We illustrate that the classical approach of combining $p$-values can be viewed as a special case of combining CDs. We also propose a weighted combination, unifying the model-based meta-analysis approaches under the proposed framework of combining CDs. To our knowledge, this is the first time that the classical methods of combining $p$-values and the model-based meta-analysis methods are placed under a unified framework. The unifying framework not only has theoretical value, but can also promote broader applications of, and easier access to, meta-analysis through development of a common computer program for a variety of approaches.

The proposed CD combining framework also allows us to develop new methodologies. Specifically, we develop two robust meta-analysis approaches, with supporting asymptotic theory. In one approach, the size of each study goes to infinity, and in the other, the number of studies tends to infinity. The proposed robust meta-analysis approaches have high breakdown points and are resistant to "bad" studies. Here, "bad" studies are those whose underlying true parameter values differ from the parameter of interest, and we assume that we do not know which studies are "bad" studies. This development removes a constraint, often implicitly imposed in current practice, requiring all studies be of the same type and with the exact same underlying parameter (or hyperparameter) values. The proposed robust approaches also have high efficiency, asymptotically. We prove an oracle property implying that, in the first setting, the robust meta-analysis estimator is asymptomatically equivalent to the theoretically most efficient point estimator in fixed-effects models, regardless of whether any studies are outlying or not. The second approach can attain $(3/\pi)^{1/2} \approx 97.7\%$ efficiency asymptotically in both fixed-effects and random-effects models when there are no outlying studies. The second approach also offers some protection against model misspecification, and it has an interesting connection to an M-estimation approach, which has not been explored before.

The CD concept has a historical connection to the classical fiducial development, and recent developments on CDs also share some common goals with fiducial inference and its extensions, including recent development on "generalized fiducial" inference and the development of belief functions under Dempster–Shafer theory; see, for example, review articles by Hannig (2009), Dempster (2008), and Martin, Zhang, and Liu (2010). In particular, Dempster's rule of recombination proposes "a universal rule to combine evidence" through combining belief functions. However, as stated in Martin, Zhang, and Liu (2010), "the belief function does not satisfy long-run frequency properties" and the expression for combining belief functions is "rather complicated." To the best of our knowledge, computational difficulty and other issues have so far limited the use of the Dempster's rule of recombination in practice. In contrast to these developments, the CD concept considered here is defined and developed strictly within the frequentist domain. There is no involvement of any new theoretical framework such as fiducial reasoning or Dempster–Shafer theory. The research on combining CDs in the present article is not an attempt to provide a universal theory of combining evidence. Rather, the focus is on providing a unifying framework that subsumes most commonly used meta-analysis approaches and developing new robust approaches for practical applications. Unlike Dempster's rule of recombination, the CD-combining approach is easy to compute and can be directly related to a vast collection of practical examples, including most approaches used in the classical and current meta-analysis practices.

The article is organized as follows. Section 2 introduces a general recipe and framework for combining CDs from independent studies. Section 3 illustrates that classical methods of combining $p$-values, as well as model-based meta-analysis approaches, can be derived as special examples under the general framework of combining CDs. Section 4 develops two general robust meta-analysis approaches under different asymptotic settings and illustrates them using fixed-effects and random-effects models. Section 5 contains two numerical examples. The first uses the intravenous lidocaine treatment data studied by Normand (1999), and the other uses the stomach ulcer data provided in Efron (1996). Section 6 provides some further remarks.

## 2. A SIMPLE RECIPE AND GENERAL FRAMEWORK OF CD COMBINATION

Suppose $H_i(\theta) = H_i(\mathbf{X}_i, \theta)$, $i = 1, \ldots, k$, are CD functions for the same parameter $\theta$ from $k$ independent samples $\mathbf{X}_i$ and the sample size of $\mathbf{X}_i$ is $n_i$. By extending the classical methods of combining $p$-values, Singh, Xie, and Strawderman (2005) proposed a general recipe for combining CD functions using a coordinate-wise monotonic function from the $k$-dimensional cube $[0, 1]^k$ to the real line $\mathbb{R} = (-\infty, +\infty)$. Specifically, let $g_c(u_1, \ldots, u_k)$ be a given continuous function on $[0, 1]^k \to \mathbb{R}$ which is monotonic (without loss of generality, say, increasing) in each coordinate. Singh, Xie, and Strawderman (2005) suggested to combine the $k$ CD functions as

$$H^{(c)}(\theta) = G_c\big\{g_c(H_1(\theta), \ldots, H_k(\theta))\big\}. \tag{2.1}$$

Here, the function $G_c$ is completely determined by the monotonic $g_c$ function: $G_c(t) = P(g_c(U_1, \ldots, U_k) \le t)$, where $U_1, \ldots, U_k$ are independent U[0, 1] random variables. When the underlying true parameter values of the $k$ individual CD functions $H_i(\theta)$ are the same, it is evident that $H^{(c)}(\theta)$ is a CD function for the parameter $\theta$. This function $H^{(c)}(\theta)$ contains information from all $k$ samples, and it is referred to as a *combined CD function*. A nice feature of the proposed CD-combining method is that it does not require any information regarding how the input CD functions, $H_i(\theta)$, are obtained, aside from the assumed independence.

Although only representing a small fraction of combining approaches covered by (2.1), a special class of the general combining recipe, specified below, plays a prominent role in unifying many meta-analysis approaches currently used in practice. In this special class, the choice of the function $g_c$ is

$$g_c(u_1, \ldots, u_k) = w_1 F_0^{-1}(u_1) + \cdots + w_k F_0^{-1}(u_k), \tag{2.2}$$

where $F_0(\cdot)$ is a given cumulative distribution function and $w_i \ge 0$, with at least one $w_i \ne 0$, are generic weights for the combination. Two types of weights are considered: (a) fixed

weights in various contexts to improve the efficiency of combination, and (b) adaptive weights (or data-based weights) to deal with unknown parameters or to obtain a robust combination. When $w_1 \equiv \cdots \equiv w_k \equiv 1$, this weighted combining recipe reduces to the subclass of equal-weight combination with

$$g_c(u_1, \ldots, u_k) = F_0^{-1}(u_1) + \cdots + F_0^{-1}(u_k). \tag{2.3}$$

Singh, Xie, and Strawderman (2005) focused on the study of Bahadur efficiency in the setting where the underlying true parameter values of the $k$ individual CD functions $H_i(\theta)$ are the same. They showed that, when $g_c(u_1, \ldots, u_k) = \mathrm{DE}^{-1}(u_1) + \cdots + \mathrm{DE}^{-1}(u_k)$, the combined CD function $H^{(c)}(\theta)$ is most efficient, in terms of Bahadur slope. Here, $\mathrm{DE}(\cdot)$ is the cumulative distribution function of the standard double exponential distribution and $k$ is bounded. Singh, Xie, and Strawderman (2005) also extended this Bahadur optimality result to an empirical-Bayes like setting discussed by Efron (1993) in which, in addition to a "current study" of interest, there are also some "past studies" that may or may not provide useful information for the current study. To achieve the goal of incorporating past information with the current study (say, Study 1) under the framework of combining CDs, they considered a special form of (2.2), in which $w_1 \equiv 1$ and $w_i$ for past studies are data-dependent adaptive weights which are equal to (or, asymptotically, tending to) either 0 or 1.

In contrast, here we study methods of combining CDs to obtain an overall conclusion from separate studies in a meta-analysis setting in which the underlying true parameters of the studies may or may not be the same. Also, since Bahadur optimality has some technical problems when dealing with nonexact inferences (as is commonly the case in meta-analysis applications), instead of focusing on Bahadur efficiency, we concentrate on providing a unifying framework for various existing meta-analysis approaches and provide Fisher-type optimality (or near optimality) results whenever possible. Weighted combining to improve Fisher-efficiency of a combination, which was not discussed in Singh, Xie, and Strawderman (2005), plays a key role in this development, especially in model-based meta-analysis approaches. As illustrated in Section 3, the classical approaches of combining $p$-values (cf., Fisher 1932 and Marden 1991) as well as the model-based meta-analysis methods (e.g., Normand 1999) are special cases of this framework of combining CDs. To our knowledge, the current development has allowed, for the first time, these seemingly unrelated methods to be studied in a unified framework.

We also extend the adaptive weighting idea of Singh, Xie, and Strawderman (2005) to a meta-analysis setting, in which up to half of the studies can be allowed to have different underlying parameter values; see Section 4.1. Besides providing a robust meta-analysis approach for a set of large studies, the development also offers an improved and much stronger theoretical result on adaptive combining than Singh, Xie, and Strawderman (2005). Furthermore, it provides a Fisher-optimal combining result for the fixed-effects models in the normal case.

Additionally, we study another choice in (2.1),

$$g_c(u_1, \ldots, u_k) = w_1 u_1 + \cdots + w_k u_k, \tag{2.4}$$

which leads to our second robust approach under the asymptotic assumption that the number of studies $k$ goes to infinity. It can be shown that the combined CD function $H^{(c)}(\theta)$ obtained by using (2.4) is related to an M-estimation approach. This new robust approach is close to Fisher-optimal in normal models, and it covers both random effects and fixed-effects models.

To facilitate the use of the CD combining approaches, we present below a lemma to the effect that the general recipe (2.1) can preserve orders of error bounds (the error is quantified in term of deviating from U[0, 1] distribution at its true parameter value). The result implies that the combined CD function can preserve the convergence rates of individual aCD functions. It also allows us to use approximations of CD functions in practice. A proof of the lemma is in Appendix B.

*Lemma 1.* Suppose $\tilde{H}_i(\theta)$ is an approximate CD function that satisfies

$$\left|P(\tilde{H}_i(\theta_0) \le t) - t\right| \le \epsilon_i \quad \text{for an } \epsilon_i > 0 \text{ and all } t \in (0, 1),\ i = 1, 2, \ldots, k. \tag{2.5}$$

Then, the combined function $\tilde{H}^{(c)}(\theta) = G_c\{g_c(\tilde{H}_1(\theta), \ldots, \tilde{H}_k(\theta))\}$ is an approximate CD function satisfying

$$\left|P\big(\tilde{H}^{(c)}(\theta_0) \le t\big) - t\right| \le \sum_{i=1}^{k} \epsilon_i \quad \text{for all } t \in (0, 1).$$

## 3. UNIFYING CLASSICAL AND MODEL–BASED META–ANALYSIS APPROACHES

### 3.1 Unifying Classical *p*-Value Combining Methods

One classical approach combines the $p$-values from individual studies. Let us start with a left-sided test $K_0 : \theta \le t_0$ versus $K_1 : \theta > t_0$ for some fixed $t_0$. Denote by $p_i$ the $p$-value from the $i$th study, $i = 1, 2, \ldots, k$. Fisher (1932) suggested a simple combining method using

$$p^{(c)} = P\left\{\chi^2_{2k} \ge -2\sum_{i=1}^{k} \log(p_i)\right\} \tag{3.1}$$

as a $p$-value for all $k$ studies, and Littell and Folks (1973) established that the combination in (3.1) is Bahadur optimal. Here, $\chi^2_{2k}$ is a $\chi^2_{2k}$-distributed random variable. Another commonly used $p$-value combining method, proposed by Stouffer et al. (1949), is

$$p^{(c)} = \Phi\left(\frac{1}{\sqrt{k}}[\Phi^{-1}(p_1) + \Phi^{-1}(p_2) + \cdots + \Phi^{-1}(p_k)]\right), \tag{3.2}$$

where $\Phi$ is the cumulative distribution function of the standard normal distribution. Additional approaches of combining $p$-values in the classical meta-analysis literature include Tippett (min), sum, and max methods (see, e.g., Marden 1991 and the references therein).

For the test $K_0 : \theta \le t_0$ versus $K_1 : \theta > t_0$, the $p$-value $p_i$ depends on the value $t_0$. When $t_0$ varies, $p_i = p_i(t_0)$ forms a function on the parameter space. This function $p_i(\cdot)$ is called a *significance function* by Fraser (1991) and is also known as a *p-value function*. Singh, Xie, and Strawderman (2005, 2007) showed that a $p$-value function is usually a CD or an aCD function. Based on the development of the $p$-value function and its connection to the CD concept, the aforementioned methods of combining $p$-values are naturally linked to the approach of combining CDs.

Table 1. A list of meta-analysis approaches unified under the proposed combining CDs framework

| | |
|---|---|
| Classical approaches of combining $p$-values (from Marden 1991) | Fisher method |
| | Stouffer (normal) method |
| | Tippett (min) method |
| | Max method |
| | Sum method |
| Model-based meta-analysis approaches (from Normand 1999, table IV) | Fixed-effects model: MLE method |
| | Fixed-effects model: Bayesian method |
| | Random-effects model: Method of moment |
| | Random-effects model: REML method |
| | Random-effects model: Bayesian method (normal prior on $\theta$ and fixed $\tau$) |



More specifically, let $p_1(s), \ldots, p_k(s)$ be the $p$-value functions in the $k$ studies. In each study, the value $p_i = p_i(t_0)$ is the $p$-value of the one-sided test $K_0 : \theta \le t_0$ versus $K_1 : \theta > t_0$. Based on the equal-weight recipe (2.3) with $F_0(t) = e^t$ for $t \le 0$ or $F_0(t) = \Phi(t)$, we can get a combined CD

$$H^{(c)}(s) = P\left[\chi^2_{2k} \ge -2\sum_{i=1}^{k} \log\{p_i(s)\}\right] \tag{3.3}$$

or

$$H^{(c)}(s) = \Phi\Big(\frac{1}{\sqrt{k}}\big[\Phi^{-1}(p_1(s)) + \Phi^{-1}(p_2(s)) + \cdots + \Phi^{-1}(p_k(s))\big]\Big), \tag{3.4}$$

respectively. The $p^{(c)}$ in (3.1) equals $H^{(c)}(t_0)$ in (3.3), and the $p^{(c)}$ in (3.2) equals $H^{(c)}(t_0)$ in (3.4). Thus, the approaches of combining $p$-value functions (CDs) in (3.3) and (3.4) subsume the approaches of combining $p$-values in (3.1) and (3.2). This conclusion can be extended to the other approaches of combining $p$-values. We have verified that all five methods of combining $p$-values investigated in Marden (1991) can be subsumed under the framework of combining CDs. For Tippett (min), max, and sum methods, the $g_c$ choices in (2.1) are $g_c(u_1, \ldots, u_k) = \min(u_1, \ldots, u_k)$ or $\max(u_1, \ldots, u_k)$ or $u_1 + \cdots + u_k$, respectively.

The extension of this argument to combining $p$-values for a right-sided test is trivial. Note that $\tilde{p}_i = 1 - p_i(t_0)$ is the $p$-value for the right-sided test $K_0 : \theta \ge t_0$ versus $K_1 : \theta < t_0$, where $p_i(s)$ is the $p$-value function defined on the corresponding left-sided tests. Let $H^{(c)}(s)$ be the combined CD function from combining these $p_i(s)$ functions. Then, $1 - H^{(c)}(t_0)$ is the combined $p$-value for the right-sided test.

For a two-sided test $K_0 : \theta = t_0$ versus $K_1 : \theta \ne t_0$, the $p$-value of the test is $\tilde{p}_i = 2\min\{p_i(t_0), 1 - p_i(t_0)\}$; see Fraser (1991) and Singh, Xie, and Strawderman (2007). In this case, instead of using the $p$-value (CD) functions $p_i(s)$, we obtain the combined $p$-value $p^{(c)}$ of the two-sided test using the *centrality functions* $c_i(s) = 2\min\{p_i(s), 1 - p_i(s)\}$. In contrast to a CD function, the centrality function $c_i(s) = 2\min\{p_i(s), 1 - p_i(s)\}$ peaks around $s = t_0$. But $c_i(t_0)$ is still U[0, 1] distributed as a function of the sample. We can prove that, using the general combining recipe (2.1) and with $g_c$ in the form of (2.3), the result of combining centrality functions

$$c^{(c)}(s) = G_c\big\{g_c(c_1(s), \ldots, c_k(s))\big\}$$

is still a centrality function. Its value at $t_0$, $c^{(c)}(t_0)$, is the combined $p$-value $p^{(c)}$ for the two-sided test.

The first half of Table 1 lists the five classical $p$-value combining approaches considered in Marden (1991). The combined $p$-values of these five approaches can all be obtained under the general CD-combining framework, with an equal-weight for each individual study in the combination.

### 3.2 Unifying Model-Based Meta-Analysis Methods

Normand (1999) and Sutton and Higgins (2008) provided excellent reviews of model-based meta-analysis in modern biostatistics applications. They summarized various meta-analysis procedures under the framework of both fixed-effects and random-effects models. We illustrate in this subsection that commonly used meta-analysis procedures based on fixed-effects and random-effects models fit into our general framework of combining CDs. In particular, we use the weighted recipe (2.2), with some chosen weights, to match the estimators from combined CDs with the estimators from model-based meta-analysis approaches. For simplicity, we use here the normal models described in Normand (1999) to illustrate the unification of model-based meta-analysis. Further remarks on asymptotically normal or $t$-distributed models and other non-normal cases are provided at the end of the section.

*3.2.1 Fixed-Effects Model.* Normand (1999) used the following model to illustrate a fixed-effects meta-analysis approach

$$Y_i \overset{\text{ind}}{\sim} \mathrm{N}(\theta, s_i^2) \quad \text{for } i = 1, 2, \ldots, k, \tag{3.5}$$

where $\theta$ is the parameter of interest, $Y_i$ is a summary statistic from the $i$th study, and $s_i^2 = \mathrm{var}(Y_i)$ is assumed known. Normand (1999) reviewed under (3.5) two meta-analysis estimators: MLE and Bayes.

Under the fixed-effects model assumption (3.5), the CD function from the $i$th study is $H_i(\theta) = \Phi((\theta - Y_i)/s_i)$. Taking $F_0(t) = \Phi(t)$ and $w_i = 1/s_i$ in (2.2), we have a combined CD function

$$H^{(c)}(\theta) = \Phi\left(\frac{1}{\sqrt{\sum_{i=1}^k w_i^2}} \sum_{i=1}^{k} w_i \frac{\theta - Y_i}{s_i}\right) = \Phi\left(\left(\sum_{i=1}^{k} \frac{1}{s_i^2}\right)^{1/2} (\theta - \hat{\theta}_c)\right), \tag{3.6}$$

where $\hat{\theta}_c = (\sum_{i=1}^k Y_i/s_i^2)/(\sum_{i=1}^k 1/s_i^2)$. Thus, we can estimate $\theta$ by a normal CD function with mean $\hat{\theta}_c = (\sum_{i=1}^k Y_i/s_i^2)/(\sum_{i=1}^k 1/s_i^2)$ and variance $(\sum_{i=1}^k 1/s_i^2)^{-1}$. These are exactly the expressions for the MLE method listed in table IV of Normand (1999).

The conventional Bayes meta-analysis estimator can also fit into our CD combining framework. Following Normand (1999), we assume that the prior distribution is $\pi(\theta) \sim \mathrm{N}(0, \sigma_0^2)$. As described in Example A.2 in Appendix A.1, the prior function $\pi(\theta) \sim \mathrm{N}(0, \sigma_0^2)$ can be viewed as a CD function, assuming that there was a prior study whose summary statistic $Y_0$ had variance $\mathrm{var}(Y_0) = \sigma_0^2$ and realization (observation) $Y_0 = 0$. Write $H_0(\theta) = \Phi((\theta - Y_0)/\sigma_0) = \Phi(\theta/\sigma_0)$, the CD function from the prior study. By taking $F_0(t) = \Phi(t)$, $w_0 = 1/\sigma_0$, and $w_i = 1/s_i$ for $i = 1, \ldots, k$, and including $H_0(\theta)$ of the prior study in the combination in (2.2), we have a combined CD function

$$H^{(c)}(\theta) = \Phi\left(\frac{1}{\sqrt{\sum_{i=0}^k w_i^2}} \sum_{i=0}^k w_i \frac{\theta - Y_i}{s_i}\right)$$

$$= \Phi\left(\left(\sum_{i=1}^k \frac{1}{s_i^2} + \frac{1}{\sigma_0^2}\right)^{1/2} (\theta - \hat{\theta}_B)\right),$$

where $\hat{\theta}_B = (\sum_{i=1}^k Y_i/s_i^2)/(\sum_{i=1}^k 1/s_i^2 + 1/\sigma_0^2)$. Thus, we estimate the parameter $\theta$ by a normal CD function with mean $\hat{\theta}_B$ and variance $(\sum_{i=1}^k 1/s_i^2 + 1/\sigma_0^2)^{-1}$. These are exactly the expressions for the Bayes method listed in table IV of Normand (1999).

*3.2.2 Random-Effects Model.* The random-effects model described in Normand (1999) is a hierarchical model:

$$Y_i|(\theta_i, s_i) \overset{\text{ind}}{\sim} \mathrm{N}(\theta_i, s_i^2) \qquad \text{and} \qquad \theta_i|\theta, \tau^2 \overset{\text{ind}}{\sim} \mathrm{N}(\theta, \tau^2)$$
$$\text{for } i = 1, 2, \ldots, k, \quad (3.7)$$

where $\theta_i$ is the study-specific mean (random effect), and $\theta$ and $\tau^2$ are hyperparameters for $\theta_i$. The variance of the given $i$th study $s_i^2$ is assumed known. From (3.7), we have $Y_i \overset{\text{ind}}{\sim} \mathrm{N}(\theta, \tau^2 + s_i^2)$. Thus, based on each study, we can construct a CD function $H_i(\theta) = \Phi((\theta - Y_i)/(\tau^2 + s_i^2)^{1/2})$. Taking $F_0(t) = \Phi(t)$ and $w_i = 1/(\tau^2 + s_i^2)^{1/2}$ in (2.2), it follows that the combined CD function is

$$H^{(c)}(\theta) = \Phi\left(\left(\sum_{i=1}^k \frac{1}{\tau^2 + s_i^2}\right)^{1/2} (\theta - \hat{\theta}_c)\right), \quad (3.8)$$

where $\hat{\theta}_c = \{\sum_{i=1}^k Y_i/(\tau^2 + s_i^2)\}/\{\sum_{i=1}^k 1/(\tau^2 + s_i^2)\}$. Therefore, the combined CD function has mean $\hat{\theta}_c$ and variance $\{\sum_{i=1}^k 1/(\tau^2 + s_i^2)\}^{-1}$.

Replacing $\tau^2$ with the DerSimonian and Laird estimator $\hat{\tau}_{\mathrm{DL}}^2$, the combined CD function $H^{(c)}(\theta)$ then has mean $\hat{\theta}_{\mathrm{DL}} = \{\sum_{i=1}^k Y_i/(\hat{\tau}_{\mathrm{DL}}^2 + s_i^2)\}/\{\sum_{i=1}^k 1/(\hat{\tau}_{\mathrm{DL}}^2 + s_i^2)\}$ and variance $\{\sum_{i=1}^k 1/(\hat{\tau}_{\mathrm{DL}}^2 + s_i^2)\}^{-1}$. These expressions match the meta-analysis estimators for the method of moments listed in table IV of Normand (1999).

If $\tau^2$ is estimated by its REML estimator $\hat{\tau}_{\mathrm{R}}^2$, the combined CD of $\theta$ has mean $\hat{\theta}_{\mathrm{R}} = \{\sum_{i=1}^k Y_i/(\hat{\tau}_{\mathrm{R}}^2 + s_i^2)\}/\{\sum_{i=1}^k 1/(\hat{\tau}_{\mathrm{R}}^2 + s_i^2)\}$ and variance $\{\sum_{i=1}^k 1/(\hat{\tau}_{\mathrm{R}}^2 + s_i^2)\}^{-1}$. These expressions match the REML type of estimators listed in table IV of Normand (1999).

The second half of Table 1 includes the model-based meta-analysis methods listed in table IV of Normand (1999). The meta-analysis estimators from these approaches can all be obtained under our CD combining framework, using the weighted recipe (2.2). Since an appropriate choice of weights can improve the combining efficiency in normal models, this explains why an approach of combining $p$-values is typically not as efficient as a model-based approach when the model assumption holds. We remark that, in the random-effects model (3.7) with $\tau^2 \neq 0$, neither $\hat{\theta}_{\mathrm{DL}}$ nor $\hat{\theta}_{\mathrm{R}}$ is a consistent estimator of $\theta$ when $k$ is bounded. The implications of this fact are further discussed in Section 4.

In the fixed-effects and random-effects models (3.5) and (3.7), the variance $s_i^2$ in each study is assumed known, following Normand (1999). In practice, the variance $s_i^2$ is often estimated. This will not change our methodology, and the same CD-combining methods still apply and lead to matches to their corresponding items in the conventional meta-analysis. The only difference in using an estimated $s_i^2$ is that $H_i(\theta)$ and $H^{(c)}(\theta)$ now are aCD instead of exact CD functions. There may be an interesting twist in this case. For example, in the fixed-effects model, we can replace the aCD function $H_i(\theta) = \Phi((\theta - Y_i)/s_i)$ with an exact CD function $H_i(\theta) = F_{t_{n_i-1}}((\theta - Y_i)/s_i)$, whenever the exact $t$-distribution statement applies. Here, $F_{t_{n_i-1}}$ is the cumulative distribution function of the $t_{n_i-1}$ distribution. Combining these $t$-CD functions probably will not yield a meta-analysis estimator that matches any conventional meta-analysis estimator from directly combining point estimators (i.e., $Y_i$'s), except asymptotically. But the estimator computed from combining these $t$-CD functions may have better performance than the MLE meta-analysis estimator corresponding to the $H^{(c)}(\theta)$ in (3.6) when $n_i$ is small or only moderately large, noting that the $t$-based approach uses exact distributions with no asymptotic approximations.

As mentioned in Section 2, the proposed CD combining method does not require any information regarding how $H_i(\theta)$ are obtained. Neither is limited to normal or asymptotically normal cases. For instance, the examples of using $p$-value functions discussed in Section 3.1 do not rely on normality or asymptotic normality. The CDs ($p$-value functions) can come from any one-sided tests under any distributions, including those from nonparametric tests where the exact form of the underlying distribution is not unavailable. Another example is to combine independent bootstrap distributions from different studies, noting that bootstrap distributions are aCD functions (Efron 1998; Singh, Xie, and Strawderman 2005). Section 5.3 of Singh, Xie, and Strawderman (2005) provided an example of combining bootstrap distributions, as a way to save computing effort in a setting involving heavy computations on a large dataset. Indeed, the CD combing framework for meta-analysis is very broad. It provides a structure and opportunity to explore and discover sensible meta-analysis approaches that are not possible, or difficult to get, under the conventional approach of combining point estimators.

## 4. ROBUST META–ANALYSIS APPROACHES

The development of robust meta-analysis methods in this section offers another illustration that the CD-combining framework can lead to new approaches. Section 4.1 considers the setting of combining large studies, and the development in Section 4.2 assumes that the number of studies goes to infinity.

### 4.1 Robust Meta-Analysis of a Set of Large Studies

Suppose there are $k$ studies of the same treatment and the sample sizes in these studies are $n_1, n_2, \ldots, n_k$, respectively. Throughout this subsection, we assume that each $n_i$ goes to infinity and $k$ is bounded. For notational simplicity, we assume that the $n_i$ go to infinity at the same rate $n$, although this assumption can be relaxed to a certain degree.

Assume that we are interested in a specific characteristic of the treatment described by a parameter $\theta$. Let the underlying true value of the parameter $\theta$ for the $i$th study be $\theta_i^{(0)}$, which may not be the same across all $k$ studies (we do not know which studies are different). The parameter of interest is defined as

$$\theta_0 = \text{median}\{\theta_1^{(0)}, \ldots, \theta_k^{(0)}\}. \tag{4.1}$$

Define a set $\mathcal{I}_0 = \{i : \theta_i^{(0)} = \theta_0\}$. If the number of studies in $\mathcal{I}_0$ is greater than $[k/2]$, $\theta_0$ is the true parameter value of the majority of the studies. Here, $[\cdot]$ is the rounding function for integers.

Let $H_i(\theta)$ be a CD function of the parameter $\theta$ obtained from the $i$th study, $i = 1, 2, \ldots, k$. In this subsection, we assume that the $\theta_i^{(0)}$ are fixed and each $H_i(\theta)$ satisfies:

*Condition (A).* For any fixed $\gamma, 0 < \gamma < \frac{1}{2}$, $L_i(\gamma) = H_i^{-1}(1-\gamma) - H_i^{-1}(\gamma) \to 0$, in probability, as $n \to \infty$.

Here, $H_i^{-1}(\beta)$ is the $\beta$-quantile of $H_i(\theta)$; i.e., it solves the equation $H_i(\theta) = \beta$. Condition (A) is the same as Condition (3.1) of Singh, Xie, and Strawderman (2007), where this condition was imposed for obtaining consistent point estimators from a CD function. This condition essentially assumes that the $i$th study can produce a consistent estimator for its underlying parameter $\theta_i^{(0)}$. Condition (A) is equivalent to condition

*Condition (A′).* For any fixed $\delta > 0$, $H_i(\theta_i^{(0)} - \delta) \to 0$ and $H_i(\theta_i^{(0)} + \delta) \to 1$, in probability, as $n \to \infty$.

A proof of the equivalence is in Appendix B. We interpret Condition (A′) as follows: as $n$ increases, the information contained in the CD function $H_i(\theta)$ becomes more and more concentrated around $\theta_i^{(0)}$.

A special example of the above set-up is the following fixed-effects model:

$$Y_i \overset{\text{ind}}{\sim} \text{N}(\theta_i, s_i^2) \quad \text{for } i = 1, 2, \ldots, k, \tag{4.2}$$

where $\theta_i$ is a study-specific parameter with true underlying value $\theta_i^{(0)}$. If $\theta_i \equiv \theta$ (with the same true underlying value $\theta_i^{(0)} \equiv \theta_0$) for all studies, model (4.2) reduces to the conventional fixed-effects model (3.5). In Model (4.2), $n$ can be a generic sample size of the order $1/s_i^2$, assuming that all $1/s_i^2$ are of the same order. In either (3.5) or (4.2), the CD function from the $i$th study is $H_i(\theta) = \Phi((\theta - Y_i)/s_i)$. Since we typically have $s_i = O_p(n^{-1/2})$, it follows that $L_i(\gamma) = 2s_i\Phi^{-1}(1-\gamma) = O_p(n^{-1/2}) \to 0$ for any $0 < \gamma < 1/2$. Thus, Condition (A) is satisfied in the fixed-effects models.

The key idea in developing our robust approach under the current setting is to obtain a set of data-dependent adaptive weights, so that, as sample sizes increase, the CD-combining procedure (2.2) tends to combine only the correct information and down weight or exclude studies containing little information about the parameter of interest.

To illustrate this idea, let us first consider an empirical-Bayes-like setting studied by Efron (1993) and Singh, Xie, and Strawderman (2005). In particular, suppose that there is a "current" study and some "past" studies; without loss of generality in our context, say, the first study is the "current" study and the remaining $k-1$ are the "past" studies. We are interested in making inference about the true parameter value $\theta_1^{(0)}$, based on the first study and incorporating information from the "past" studies. Denote by a set $\mathcal{I}_{1,0} = \{i : \theta_i^{(0)} = \theta_1^{(0)}\}$. Although we do not know the membership of $\mathcal{I}_{1,0}$ except for the first study, we would like to combine all studies in $\mathcal{I}_{1,0}$, excluding or down weighing all studies outside of $\mathcal{I}_{1,0}$. This task may be achieved asymptotically by using a set of adaptive weights $w_i^{(a)}$ in the weighted combination (2.2) such that

$$\lim_{n\to\infty} w_i^{(a)} = \begin{cases} 1 & \text{if } \theta_i^{(0)} = \theta_1^{(0)} \\ 0 & \text{if } \theta_i^{(0)} \neq \theta_1^{(0)} \end{cases} \quad \text{for } i = 1, 2, \ldots, k. \tag{4.3}$$

One set of adaptive weights $w_i^{(a)}$ that satisfies (4.3) is the following. Let $\hat{\theta}_i$ be a consistent estimator of $\theta_i^{(0)}$ from the $i$th study and $K(t)$ be a symmetric kernel function, $\int K(t)\,dt = 1$, $\int tK(t)\,dt = 0$, and $\int t^2K(t)\,dt < \infty$. Also, let $b_n$ be a tuning constant such that $b_n \to 0$ and $|\hat{\theta}_i - \theta_i^{(0)}| = o_p(b_n)$. We define

$$w_i^{(a)} = K\left(\frac{\hat{\theta}_1 - \hat{\theta}_i}{b_n}\right)\Big/K(0) \quad \text{for } i = 1, 2, \ldots, k. \tag{4.4}$$

One example that we use in our numerical study in Section 5 is $K(t) = 2\min\{\Phi(t), 1 - \Phi(t)\}$, $\hat{\theta}_i = H_i^{-1}(1/2)$, and $b_n \propto \{H_1^{-1}(3/4) - H_1^{-1}(1/4)\}^{1/2}$. Here, under Condition (A), $\hat{\theta}_i = H_i^{-1}(1/2)$ is a median-unbiased consistent estimator of $\theta_i^{(0)}$ with the same convergence rate as $L_i(\delta)$; cf., theorem 3.1 of Singh, Xie, and Strawderman (2007). In the fixed-effects model (4.2), this choice leads to $|\hat{\theta}_i - \theta_i^{(0)}| = O_p(s_i) = O_p(n^{-1/2})$ and $b_n \propto O_p(s_i^{1/2}) = O_p(n^{-1/4})$; thus, it ensures (4.3).

Let $H_1^{(c)}(\theta)$ be the combined CD function of $H_1(\theta), \ldots, H_k(\theta)$, using the weighted recipe (2.2) where the weights $w_i = w_i^{(a)}w_i^{(e)}$ with the adaptive weights $w_i^{(a)}$ in (4.4) and also possibly other given weights $w_i^{(e)}$ from efficiency considerations. (For instance, $w_i^{(e)} = 1/s_i$ in the normal model considered later in Theorem 2.) Let $H_{1,0}^{(c)}(\theta)$ be the the corresponding target in the ideal case, that combines all studies in $\mathcal{I}_{1,0}$ and excludes all studies outside of $\mathcal{I}_{1,0}$. That is, $H_{1,0}^{(c)}(\theta)$ is obtained in the same way as $H_1^{(c)}(\theta)$, but the adaptive weight $w_i^{(a)}$ in the combination is replaced by 1 if $\theta_i^{(0)} = \theta_1^{(0)}$ and by 0 if $\theta_i^{(0)} \neq \theta_1^{(0)}$. We use the following lemma in the development. A proof is given in Appendix B.

*Lemma 2.* Consider the settings described above. Suppose $F_0$ in (2.2) is selected such that

$$K(1/b_n)F_0^{-1}\big(H_i\big(\theta_1^{(0)}\big)\big) \to 0 \quad \text{in probability} \tag{4.5}$$

for all $1 \le i \le k$. We have

(i) $\sup_\theta |H_1^{(c)}(\theta) - H_{1,0}^{(c)}(\theta)| \to 0$ in probability, and $H_1^{(c)}(\theta)$ is an aCD for the parameter $\theta_1^{(0)}$.

(ii) The median of the combined CD function $\hat\theta_1^{(c)} = {H_1^{(c)}}^{-1}(1/2) \to \theta_1^{(0)}$ in probability, as $n \to \infty$.

The first result in Lemma 2 is a significantly improved version of theorem 3.5 of Singh, Xie, and Strawderman (2005), with a much stronger result under weaker conditions. Also, Singh, Xie, and Strawderman (2005) only covered the special case with $w_i^{(e)} \equiv 1$.

A set of sufficient conditions for (4.5) is as follows. Suppose that $b_n = O_p((\log n)^{-2})$ and, as $|t| \to \infty$, $K(t) \to 0$ exponentially fast. Let the tail convergence rate in Condition (A′) be such that

$$\max\big[\log\big\{H_i\big(\theta_i^{(0)} - \delta\big)\big\}, \log\big\{1 - H_i\big(\theta_i^{(0)} + \delta\big)\big\}\big]/n^s \to 0, \quad \text{as } n \to \infty, \text{ for an } s > 0. \tag{4.6}$$

Then, for any $F_0$ such that $\min\{F_0(t), 1 - F_0(t)\} \to 0$ exponentially fast, as $|t| \to \infty$, condition (4.5) is satisfied. Most CD functions in the exponential family satisfy (4.6).

We are now ready to propose our robust meta-analysis estimator of $\theta_0$, the parameter of interest defined in (4.1). For each $i$, $i = 1, 2, \ldots, k$, we obtain a combined CD function $H_i^{(c)}(\theta)$ for its true parameter value $\theta_i^{(0)}$. As claimed in Lemma 2, the median of the $i$th combined CD function $\hat\theta_i^{(c)} = {H_i^{(c)}}^{-1}(1/2)$ is a consistent estimator of $\theta_i^{(0)}$. A robust meta-analysis estimator of $\theta_0$ is then

$$\hat\theta^{(c)} = \text{median}\big\{\hat\theta_1^{(c)}, \ldots, \hat\theta_k^{(c)}\big\}.$$

Denote by $H^{(o)}(\theta)$ the combined CD function that corresponds to $\hat\theta^{(c)}$. Specifically, when $k = 2m + 1$ is odd, $H^{(o)}(\theta)$ is the CD function that corresponds to the $m$th smallest $\hat\theta_{(m)}^{(c)}$; when $k = 2m$ is even, $H^{(o)}(\theta)$ is the average of the two CD functions that correspond to $\hat\theta_{(m)}^{(c)}$ and $\hat\theta_{(m+1)}^{(c)}$.

Let $H_0^{(c)}(\theta)$ be the corresponding combined CD function, in the ideal case, using only the studies in the set $\mathcal{I}_0 = \{i : \theta_i^{(0)} = \theta_0\}$, assuming $\mathcal{I}_0 \ne \emptyset$ and the membership of $\mathcal{I}_0$ is known. Theorem 1 below suggests that $\hat\theta^{(c)}$ is a robust and consistent estimator of $\theta_0$. Also, when $\mathcal{I}_0 \ne \emptyset$ and $H_0^{(c)}(\theta)$ is well defined, $H^{(o)}(\theta)$ is asymptotically the same as $H_0^{(c)}(\theta)$. A proof is given in Appendix B.

*Theorem 1.* Under the setting of Lemma 2, we have, as $n \to \infty$:

(i) The breakdown point of the estimator $\hat\theta^{(c)}$ is $[k/2]/k$.

(ii) The estimator $\hat\theta^{(c)} \to \theta_0$, in probability.

If further $\mathcal{I}_0 \ne \emptyset$, we have, as $n \to \infty$,

(iii) $\sup_\theta |H^{(o)}(\theta) - H_0^{(c)}(\theta)| \to 0$, in probability, and $H^{(o)}(\theta)$ is an aCD for $\theta_0$.

In the special case of fixed-effects model (4.2), using the combining recipe (2.2) with $F_0 = \Phi$ and $w_i = w_i^{(a)}/s_i$, we have for the first study

$$H_1^{(c)}(\theta) = \Phi\left(\left(\sum_{i=1}^k \frac{w_i^{(a)}}{s_i^2}\right)^{1/2}\big(\theta - \hat\theta_1^{(c)}\big)\right)$$

and

$$H_{1,0}^{(c)}(\theta) = \Phi\left(\left(\sum_{i\in\mathcal{I}_{1,0}} \frac{1}{s_i^2}\right)^{1/2}\big(\theta - \tilde\theta_{1,0}^{(c)}\big)\right),$$

where $\hat\theta_1^{(c)} = (\sum_{i=1}^k w_i^{(a)} Y_i/s_i^2)/(\sum_{i=1}^k w_i^{(a)}/s_i^2)$ and $\tilde\theta_{1,0}^{(c)} = (\sum_{i\in\mathcal{I}_{1,0}} Y_i/s_i^2)/(\sum_{i\in\mathcal{I}_{1,0}} 1/s_i^2)$. In this case $\tilde\theta_{1,0}^{(c)}$ is the Fisher-optimal "estimator" of $\theta_1^{(0)}$ when the membership of $\mathcal{I}_{1,0}$ is known. Similar formulas can be obtained for the other studies. In addition, from the expression of $H_{1,0}^{(c)}(\theta)$, we can see that the optimal CD function for $\theta_0$ from all studies in $\mathcal{I}_0 = \{i : \theta_i^{(0)} = \theta_0\}$ is

$$H_0^{(c)}(\theta) = \Phi\left(\left(\sum_{i\in\mathcal{I}_0} \frac{1}{s_i^2}\right)^{1/2}\big(\theta - \tilde\theta_0^{(c)}\big)\right)$$

assuming $\mathcal{I}_0 \ne \emptyset$ and the membership of $\mathcal{I}_0$ is known. Here, $\tilde\theta_0^{(c)} = (\sum_{i\in\mathcal{I}_0} Y_i/s_i^2)/(\sum_{i\in\mathcal{I}_0} 1/s_i^2)$ is the most efficient meta-analysis "estimator" when $\mathcal{I}_0$ is known. Typically, $\tilde\theta_0^{(c)} - \theta_0 = O_p(n^{-1/2})$.

In this special case of fixed-effects model (4.2), in addition to Theorem 1, we have a stronger oracle result. In particular, Theorem 2 below states that, under the fixed-effects model (4.2), the proposed robust estimator $\hat\theta^{(c)}$ without knowledge of $\mathcal{I}_0$ (except assuming $\mathcal{I}_0 \ne \emptyset$) is asymptotically equivalent to $\tilde\theta_0^{(c)}$, up to the rate of $o_p(n^{-1/2})$. Thus, $\hat\theta^{(c)}$ is an asymptotically efficient estimator of $\theta_0$, regardless of whether there are any outlying studies or not. A proof of the theorem is in Appendix B.

*Theorem 2.* Under the fixed-effects model (4.2) and when $\mathcal{I}_0 \ne \emptyset$, we have, as $n \to \infty$, $n^{1/2}|\hat\theta^{(c)} - \tilde\theta_0^{(c)}| \to 0$. Thus, $\hat\theta^{(c)}$ is also an asymptotically efficient estimator of $\theta_0$.

### 4.2 Robust Meta-Analysis of a Large Number of Studies

Section 4.1 does not cover conventional random-effects models. It can be verified that, under a random-effects model such as (3.7), the underlying population parameter $\theta$ cannot be consistently estimated without requiring the number of studies $k$ to tend to infinity, even if the sample sizes $n_i$ of all $k$ studies go to infinity. To expand our development to cover random-effects models, we develop in this subsection a general robust meta-analysis assuming that the number of studies goes to infinity. Unless specifically stated otherwise, the sample sizes in the studies can either be bounded or tend to infinity.

Suppose we have a large number $k$ of independent studies. Along the lines of the formulation in Huber (1964), we assume that the true parameters of the studies come from a contaminated distribution

$$\theta_i \sim (1 - \epsilon)D_0(\theta) + \epsilon D_\epsilon(\theta), \qquad i = 1, 2, \ldots, k, \tag{4.7}$$

where $D_0$ and $D_\epsilon$ are distribution functions of good and contaminating populations, respectively, and $\epsilon$, $0 \le \epsilon < 1/2$, is an unknown mixing parameter. Denote by $\theta_0$ and $\theta_*$ the population mean of the good population $D_0$ and the contaminated population $(1-\epsilon)D_0 + \epsilon D_\epsilon$, respectively. The parameters $\theta_0 \equiv \theta_*$ when $\epsilon \equiv 0$ or the contamination distribution $D_\epsilon$ is symmetric around $\theta_0$.

A special example of (4.7) is a random-effects model,

$$\begin{aligned} Y_i|\theta_i, s_i &\overset{\text{ind}}{\sim} \mathrm{N}(\theta_i, s_i^2) \qquad \text{and} \\ \theta_i|(\theta, \tau^2) &\overset{\text{ind}}{\sim} (1-\epsilon)\mathrm{N}(\theta, \tau^2) + \epsilon D_\epsilon(\theta) \end{aligned} \tag{4.8}$$

for an unknown contaminating population $D_\epsilon$. In absence of contamination (i.e., $\epsilon \equiv 0$), this model is the same as the standard random-effects model (3.7). Furthermore, model (4.7) also covers fixed-effects models. Specifically, for a given $k$, the fixed-effects model (4.2), can be be treated as a special case of (4.7) with $D_0$ being the Dirac delta function at $\theta_0$ and $D_\epsilon$ being a combination of Dirac delta functions at $\theta_i^{(0)}$ when $i \in \overline{\mathcal{I}}_0 = \{i : \theta_i^{(0)} \ne \theta_0\}$.

Let $H_i(\theta)$ be a CD function for $\theta_*$ based on the sample from the $i$th study. Our robust proposal is to use the special weighted combination (2.4). Since the $k$ studies are independent and a U[0, 1] random variable has mean $1/2$ and variance $1/12$, it leads to a combined aCD function (as $k \to \infty$),

$$H^{(c)}(\theta) = \Phi\left( k^{-1/2} \sum_{i=1}^{k} w_i \left\{ H_i(\theta) - \frac{1}{2} \right\} \Big/ s_c \right), \tag{4.9}$$

where $s_c^2 = k^{-1} \sum_{i=1}^k w_i^2 \mathrm{E}\{H_i(\theta_*) - 1/2\}^2 = (12k)^{-1} \sum_{i=1}^k w_i^2$. We remark that, in practice, we typically do not need a very large $k$ to have a good asymptotic performance from $H^{(c)}(\theta)$ in (4.9), noting that the sum of $k$ U[0, 1]-distributed random variables can approximate a normal distribution fairly well even when $k$ is quite small.

In the combined CD function $H^{(c)}(\theta)$ in (4.9), the impact of each study is bounded. In fact, the inference based on $H^{(c)}(\theta)$ in (4.9) is asymptotically equivalent to the inference of an M-estimation approach that solves the following estimating equation:

$$\sum_{i=1}^{k} w_i \left\{ H_i(\theta) - \frac{1}{2} \right\} = 0. \tag{4.10}$$

Note that, since $\mathrm{E}\{H_i(\theta_*)\} = 1/2$, $\theta_*$ is also the solution to the equation $\sum_{i=1}^k w_i \mathrm{E}\{H_i(\theta) - 1/2\} = 0$.

To simplify our theoretical discussion, we assume that the weights $w_i$ considered here are fixed and nonadaptive weights, although under some mild restrictions the development can be extended to include adaptive weights including those used in Section 4.1 or something similar to those used in Hu and Zidek (2002) or Wang and Zidek (2005) in the context of weighted likelihood. The theoretical developments, except for the last result on Fisher efficiency, hold for any fixed weights. For normal or normal-like CDs, we suggest using $1/v_i$ with $v_i$ being a measure of the scale of $H_i(\theta)$; see a later example in (4.12) and Theorem 4 in which $v_i$ is the standard deviation of the normal CD $H_i(\theta) = \Phi((\theta - Y_i)/v_i)$. However, if achieving a high breakdown point is a key concern, choosing $w_1 = \cdots = w_k = 1$ guarantees an (asymptotic) breakdown point of $1/2$, as shown in Theorem 2 below. In general, we borrow a line from weighted likelihood developments: "The best choice of weights will depend on context" (Hu and Zidek 2002).

The following theorem implies that the median of $H^{(c)}(\theta)$, $\hat{\theta}^{(c)} = {H^{(c)}}^{-1}(1/2)$, is a robust and consistent estimator of $\theta_*$. A proof of the theorem is in Appendix B.

*Theorem 3.* Under the setting above, let $w_{(1)}, w_{(2)}, \ldots, w_{(k)}$ be the ordered weights (in decreasing order) in the weighted combination (4.9).

(i) As $k \to \infty$, $\hat{\theta}^{(c)} = {H^{(c)}}^{-1}(1/2)$ is a consistent estimator of $\theta_*$.

(ii) The breakdown point of $\hat{\theta}^{(c)} = {H^{(c)}}^{-1}(1/2)$ is

$$\min_{1 \le t \le k/2} \left\{ t : \sum_{i=1}^{t} w_{(i)} \ge \sum_{i=t+1}^{k} w_{(i)} \right\} \Big/ k.$$

When the weights are all the same, the breakdown point is $1/2$.

The above combining approach can be extended to combine CD-like functions. Here, a CD-like function is a function that follows Definition A.1 in Appendix A.1 except that the U[0, 1] distribution assumption is not required. In this case, we replace $H^{(c)}(\theta)$ in (4.9) by

$$\tilde{H}^{(c)}(\theta) = \Phi\left( k^{-1/2} \sum_{i=1}^{k} w_i \left\{ H_i(\theta) - \frac{1}{2} \right\} \Big/ \hat{s}_c \right), \tag{4.11}$$

where $\hat{s}_c^2 = k^{-1} \sum_{i=1}^k w_i^2 \{H_i(\hat{\theta}^{(c)}) - 1/2\}^2$ and $\hat{\theta}^{(c)} = {H^{(c)}}^{-1}(1/2)$ is the solution of Equation (4.10). We can prove that this $\tilde{H}^{(c)}(\theta)$ is still an aCD for $\theta_*$, even if the input $H_i(\theta)$'s are only CD-like functions. Here, $\theta_*$ is the solution to equation $\sum_{i=1}^k w_i \mathrm{E}\{H_i(\theta) - 1/2\} = 0$. We can also prove that the results of Theorem 3 still hold, where $\hat{\theta}^{(c)} = {H^{(c)}}^{-1}(1/2)$ is also the median of $\tilde{H}^{(c)}(\theta)$. When all input $H_i(\theta)$'s are indeed CD functions, $\hat{s}_c^2 \to s_c^2$ and $\tilde{H}^{(c)}(\theta)$ is asymptotically the same as $H^{(c)}(\theta)$ in (4.9).

This extension to $\tilde{H}^{(c)}(\theta)$ provides some protection from model misspecification. For instance, we may wrongly assume that model (3.7) is true when in fact the true model is the contaminated model (4.8). In this case, the function $H_i(\theta) = \Phi((\theta - Y_i)/(s_i^2 + \tau^2)^{1/2})$, a CD function under model (3.7), is not a CD function for either $\theta_0$ or $\theta_*$ under model (4.8). Thus, combining such $H_i(\theta)$ functions amounts to combining CD-like functions. In this case, $\hat{s}_c^2$ is a correct variance estimator, and the use of $\tilde{H}^{(c)}(\theta)$, instead of $H^{(c)}(\theta)$, provides a protection against model misspecification. The same statement also applies when the true model is a random-effects model but we wrongly apply a fixed-effects model. In addition, when the task of constructing a CD or aCD function from each study is difficult, for instance, under the contaminated model with small studies, the extension of combining CD-like functions is also practically useful.

Finally, we examine the efficiency of this robust combination under the standard fixed effects model (3.5) and the standard random-effects model (3.7), assuming there are no outlying studies. In this case, a CD function from the $i$th study is $H_i(\theta) = \Phi((\theta - Y_i)/v_i)$, where $v_i^2 = s_i^2$ in the fixed-effects model (3.5) and $v_i^2 = s_i^2 + \tau^2$ in the random-effects model (3.7).

Table 2. Meta-analysis results for effect of lidocaine on mortality

| | Conventional approach | | Robust approach | |
|---|---|---|---|---|
| | Original data | Contaminated data | Original data | Contaminated data |
| Estimate | 0.0294 | 0.0174 | 0.0290 | 0.0287 |
| sd | 0.0131 | 0.0136 | 0.0131 | 0.0138 |
| 95% CI | (0.0038, 0.0549) | (−0.0091, 0.0439) | (0.0033, 0.0546) | (0.0016, 0.0556) |
| Conclusion | $\theta_0 > 0$ | $\theta_0 = 0$ (not significant) | $\theta_0 > 0$ | $\theta_0 > 0$ |

When we pick $w_i = 1/v_i$, we have a combined normal aCD function (as $k \to \infty$),

$$H^{(c)}(\theta) = \Phi\left(\sqrt{12}\sum_{i=1}^{k}(1/v_i)\left\{\Phi\left(\frac{\theta-\hat{\theta}_i}{v_i}\right)-\frac{1}{2}\right\}\Big/\left(\sum_{i=1}^{k}1/v_i^2\right)^{1/2}\right). \quad (4.12)$$

From the standard asymptotic argument, we also know that the asymptotically efficient meta-analysis estimator is $\tilde{\theta}_0^{(c)} = \sum_{i=1}^{k}(Y_i/v_i^2)/(\sum_{j=1}^{k}1/v_j^2)$. The aCD function corresponding to $\tilde{\theta}_0^{(c)}$ is

$$H_0^{(c)}(\theta) = \Phi\big((\theta-\tilde{\theta}_0^{(c)})/\tilde{s}_c\big),$$

where $\tilde{s}_c^2 = 1/\sum_{i=1}^{k}(1/v_i^2)$. But, clearly, this optimal estimator $\tilde{\theta}_0^{(c)}$, thus $H_0^{(c)}(\theta)$, lacks robustness, with their breakdown points equal to 0 in the limit.

Theorem 4 below compares the efficiency of $H^{(c)}(\theta)$ with $H_0^{(c)}(\theta)$, in the sense of the ratio of the lengths of the confidence intervals for $\theta_0$ at the same confidence level. For notational simplicity, we assume that $n_i \propto 1/s_i^2 \to \infty$ at the same rate, say $n$. A proof of Theorem 4 is provided in Appendix B.

*Theorem 4.* Under the standard fixed-effects model (3.5) or the standard random-effects model (3.7), as $k \to \infty$ and $n \to \infty$, the asymptotic relative efficiency of $H^{(c)}(\cdot)$ compared to $H_0^{(c)}(\cdot)$ is $(3/\pi)^{1/2} \approx 0.977$.

## 5. NUMERICAL STUDIES

We perform numerical studies to examine the proposed robust meta-analysis approaches using data from the literature on prophylactic use of lidocaine after a heart attack (Normand 1999) and on a surgical treatment for stomach ulcers (Efron 1996). In the studies, the conventional model-based meta-analysis approaches are compared with the proposed robust meta-analysis approaches.

### 5.1 Mortality Data of Prophylactic Lidocaine Use in Six Large Studies

Table 1 of Normand (1999) contained mortality data for control and intravenous lidocaine treatment from $k = 6$ studies. The sample sizes of these six studies range from 82 to 300 heart attack patients. The main parameter of interest is the difference of mortality risk between control and treatment $\theta$. Normand (1999) provided detailed statistical analysis using both fixed-effects and random-effects models. Since $k = 6$ is relatively small, we only consider fixed-effects models. The first column in Table 2 contains the results reported in Normand (1999) using a fixed-effects model. The density function of $N(0.0294, 0.0131^2)$, which is also the CD density function based on the conventional meta-analysis estimator, is plotted in Figure 1(a) as a solid curve. It is concluded from the meta-analysis that $\theta_0 > 0$ and there is a "detrimental effect" of lidocaine on the mortality rate (Normand 1999).

To illustrate the impact of potential outliers, we create an outlying study by replacing the first study data $\{39, 43, 2, 1\}$ with $\{39, 43, 2, 21\}$. That is, the one death in the control group of the first study is "mistakenly" replaced by a typographic error 21. Reported in the second column of Table 2 and the dashed density curve in Figure 1(a) are results from the same conventional meta-analysis approach but using the contaminated data. Clearly, the results change a lot, including changing the conclusion to no significant effect. The conventional fixed-effects model meta-analysis is not robust.

We reanalyze both the original, as well as the contaminated, data using the robust meta-analysis method proposed in Section 4.1. Reported in the third and fourth columns of Table 2 and the density curves in Figure 2(b) are results from the robust

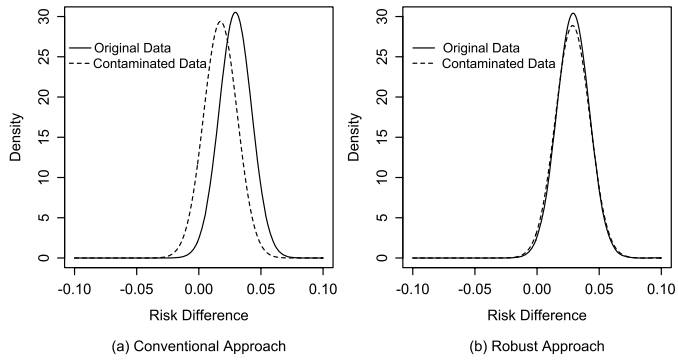


Figure 1. Comparison of the meta-analysis results of conventional method versus the robust method developed in Section 4.1. The curves for the conventional method are the corresponding normal density curves with mean and standard deviation reported in Table 2. The curves for the robust method are the CD density curves of the CD function $H^{(o)}(\theta)$ as defined in Section 4.1. For the original data, the curves are solid. For the contaminated data, the curves are broken. In the case of no outlying studies, the CD density curve [solid curve in (b)] is almost identical to the normal density curve of the corresponding conventional method [solid curve in (a)]. In the presence of one outlying study, the curve from the conventional method has a large shift, but the curve from the proposed robust method has essentially the same location.

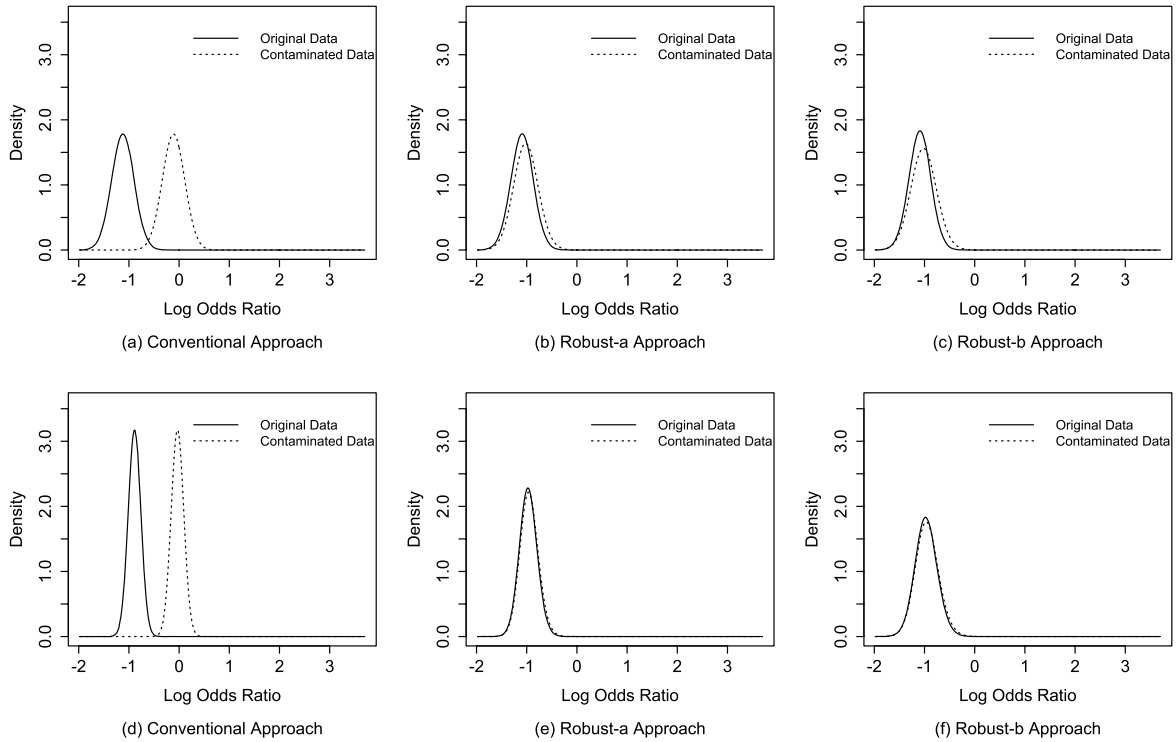


Figure 2. Comparison of the meta-analysis results from conventional method versus the robust method developed in Section 4.2. The first row [(a)–(c)] uses random-effects models, and the second row [(d)–(f)] uses fixed-effects models. The curves for the conventional methods are the normal density curves with mean and standard deviation estimates from the respective meta-analyses. The curves for the robust meta-analysis approaches are the CD density curves of the CD functions $H^{(c)}(\theta)$ and $\tilde{H}^{(c)}(\theta)$ defined in Section 4.2. Robust-a refers to $H^{(c)}(\theta)$, and Robust-b refers to $\tilde{H}^{(c)}(\theta)$. For the original data, the curves are solid lines. For the contaminated data, the curves are broken lines. In the presence of outlying studies, the curve from the conventional meta-analysis has a large shift, but the curves from the proposed robust meta-analyses have essentially the same location.

meta-analysis approach using the original and contaminated data, respectively. For the original dataset, there appears to be almost no difference between the robust and conventional fixed effects analysis. This may imply that there is no potential outlier among the six studies. But when an outlying study is injected into the original dataset, the results change only slightly from the robust approach. Clearly, the robust meta-analysis method provides a means to protect against the outlying study.

### 5.2 Stomach Ulcer Data From 41 Studies

Table 1 of Efron (1996) lists data from $k = 41$ randomized clinical trials on a new treatment for stomach ulcers from 1980 to 1989. The parameter of interest, $\theta$, is the log odds-ratio in favor of the treatment. Our goal is to obtain an overall estimate of $\theta$ using the 41 clinical trials. For the $i$th trial, we can calculate an estimate $\hat{\theta}_i$. As in Efron (1993), to obtain meaningful estimates of $\theta$, nine entries of zero are changed to 0.5 in the subsequent analysis; see Sweeting, Sutton, and Lambert (2004) for a discussion on the impact of the addition of 0.5. Also, as in Efron (1993) and others, the log odds ratios are approximated by normal distributions; see also Example A.1 in Appendix A.1.

We use both the conventional model-based meta-analysis approaches and the robust meta-analysis approaches developed in Section 4.2. The analysis is performed under random-effects models and then repeated under fixed-effects models.

The left half of Table 3 Part I and the solid curves in Figure 2(a)–(c) present results for the random-effects models. In the table, "robust-a" refers to the method using $H^{(c)}(\theta)$ in (4.9), and "robust-b" refers to the method using $\tilde{H}^{(c)}(\theta)$ in (4.11). Apparently, for this dataset there is little difference between the conventional random-effects meta-analysis and the corresponding robust meta-analysis method, using either $H^{(c)}(\theta)$ or $\tilde{H}^{(c)}(\theta)$. This seems to imply that there is little impact of outlying studies, if there are any.

To illustrate potential impact of gross outlying studies, we create a "contaminated" dataset by altering the values of the six studies whose log odds ratios are greater than 0.5. More specifically, we increase the observed log odds ratios of these six studies by a factor of 10. The contaminated data are analyzed using the same methods. The results are reported in the right half of Table 3 Part I and the broken curves in Figure 2(a)–(c). Clearly, the conventional approach is sensitive to the impact of the outlying studies. On the other hand, the results from the robust meta-analysis hardly change, indicating high resistance to the impact of outlying studies.

We also repeat the same analysis using fixed-effects models. The results are reported in Table 3 Part II and Figure 2(d)–(f). Again, the conventional approach is sensitive to the impact of the outlying studies, and the robust meta-analysis method demonstrates high resistance to the outlying studies. As expected in the conventional meta-analysis, the fixed-effects model produces narrower confidence intervals (Table 3) and more peaked CD functions (Figure 2) than the random-effects model. This is consistent with what is reported by Normand

Table 3. Meta-analysis results for log-odds-ratio of the treatment on stomach ulcer

| | Original data | | | Contaminated data | | |
|---|---|---|---|---|---|---|
| | Conventional | Robust-a | Robust-b | Conventional | Robust-a | Robust-b |
| | | | Part I. Random-effects model | | | |
| Estimate | −1.1208 | −1.0924 | −1.0924 | −0.1120 | −1.0175 | −1.0175 |
| 95% CI | (−1.5595, −0.6821) | (−1.5388, −0.6545) | (−1.5273, −0.6655) | (−0.5506, 0.3266) | (−1.4942, −0.5264) | (−1.5178, −0.5013) |
| Conclusion | $\theta_0 < 0$ | $\theta_0 < 0$ | $\theta_0 < 0$ | $\theta_0 = 0$ (not significant) | $\theta_0 < 0$ | $\theta_0 < 0$ |
| | | | Part II. Fixed-effects model | | | |
| Estimate | −0.8876 | −0.9708 | −0.9708 | −0.0325 | −0.9551 | −0.9551 |
| 95% CI | (−1.1337, −0.6415) | (−1.3097, −0.6125) | (−1.3947, −0.5180) | (−0.2786, 0.2136) | (−1.3003, −0.5799) | (−1.3910, −0.4687) |
| Conclusion | $\theta_0 < 0$ | $\theta_0 < 0$ | $\theta_0 < 0$ | $\theta_0 = 0$ (not significant) | $\theta_0 < 0$ | $\theta_0 < 0$ |



(1999) and many others. Interestingly, the combined CDs using $\tilde{H}^{(c)}(\theta)$ are notably different from those using $H^{(c)}(\theta)$ under the fixed-effects model. This difference suggests that the empirical $\hat{s}_c^2$ is different from the $s_c^2$ computed using U[0, 1] distribution. The input function $H_i(\theta) = \Phi((\theta - Y_i)/s_i)$ in the $i$th study is a CD function (with the U[0, 1]-distributed property) only when the assumed fixed-effects model (3.5) is true. Hence, this discrepancy seems to imply that this fixed-effects model may not be appropriate. Nevertheless, the combined CDs under the fixed-effects model in Figure 2(f) are not too far from those under the random-effects model in Figure 2(c), suggesting that the method based on $\tilde{H}^{(c)}(\theta)$ may offer some resistance to model misspecification.

## 6. DISCUSSION AND FURTHER REMARKS

This article develops a general framework for combining CD functions as a means to perform meta-analysis, and it provides a unifying platform that subsumes both the classical approaches of combining $p$-values and the model-based meta-analysis approaches. This unification not only can theoretically and conceptually help us understand seemingly different meta-analysis approaches, it also has practical value. An R-package *gmeta* is developed by Yang and Xie (2010) to implement this unifying framework for meta-analysis. The *gmeta* program mimics the structure of the glm() function in R, which unifies generalized linear models. The glm() function has options of "family" with different "link" functions. The gmeta() function has options of meta-analysis "method" with different choices of the monotonic function "$g_c$" (or "F0" and "weights").

The general framework also allows us to propose robust approaches for meta-analysis in line with the robustness literature in statistics. Specifically, the adaptively weighted robust method resembles adaptively weighted likelihood inference, and the robust method developed for combining a large number of studies is closely associated with M-estimation approaches. However, the CD approaches are also different from standard robust methods. For instance, the M-estimation corresponding to (4.12) has a special normal-CD-induced $\psi$-function $\psi(t) = \Phi(t) - 1/2$ that does not involve choice of any constant and has higher efficiency than that using the standard Huber $\psi$-function. The development provides a systematic, formal, and effective tool to cope with gross outlying studies in meta-analysis. Although our development is undertaken under independent settings without including covariates due to space limitation, the concepts and approaches can be extended to regression as well as more-complex meta-analysis settings, to be discussed elsewhere.

## APPENDIX A: A REVIEW OF CONFIDENCE DISTRIBUTION (CD)

### A.1 A Formal Definition of CD Function

The following CD definition is formulated in Schweder and Hjort (2002) and Singh, Xie, and Strawderman (2005). Suppose $X_1, \ldots, X_n$ are $n$ independent random draws from a population $F$ and $\mathcal{X}$ is the sample space corresponding to the data $\mathbf{X}_n = (X_1, \ldots, X_n)^T$. Let $\theta$ be a parameter of interest associated with $F$ ($F$ may contain some nuisance parameters), and let $\Theta$ be the parameter space for $\theta$.

*Definition A.1.* A function $H_n(\cdot) = H_n(\mathbf{X}_n, \cdot)$ on $\mathcal{X} \times \Theta \to [0, 1]$ is called a confidence distribution (CD) for a parameter $\theta$, if it satisfies the following two requirements: (R1) For each given $\mathbf{X}_n \in \mathcal{X}$, $H_n(\cdot)$ is a continuous cumulative distribution function; (R2) At the true parameter value $\theta = \theta_0$, $H_n(\theta_0) \equiv H_n(\mathbf{X}_n, \theta_0)$, as a function of the sample $\mathbf{X}_n$, follows the uniform distribution U[0, 1].

The function $H_n(\cdot)$ is called an asymptotic confidence distribution (aCD), if requirement (R2) above is replaced by (R2)′: At $\theta = \theta_0$, $H_n(\theta_0) \xrightarrow{W} \text{U}[0, 1]$, as $n \to \infty$, and the continuity requirement on $H_n(\cdot)$ is dropped. Here $\xrightarrow{W}$ stands for weak convergence.

We call, when it exists, $h_n(\theta) = H_n'(\theta)$ a CD density, also known as confidence density.

In nontechnical terms, a CD is a function of both the parameter and the random sample, with two requirements. The first requirement (R1) is simply that, for each given sample, a CD should be a distribution function on the parameter space. The second requirement (R2) requires that the CD function contains "balanced" (or "right") information about the true parameter value $\theta_0$ for making correct inferences. When $\theta_0$ is the true value, (R2) implies $H_n(\theta_0) \overset{\text{sto}}{=} 1 - H_n(\theta_0)$, but $H_n(\theta) \overset{\text{sto}}{\leq} 1 - H_n(\theta)$ for $\theta < \theta_0$ and $1 - H_n(\theta) \overset{\text{sto}}{\leq} H_n(\theta)$ for $\theta > \theta_0$ (see Singh, Xie, and Strawderman 2005). Here, "sto" means stochastic comparison between two random variables; for example, $Y_1 \overset{\text{sto}}{\leq} Y_2$ means $P(Y_1 \leq t) \geq P(Y_2 \leq t)$ for all $t$. We interpret this stochastic balancing equality at the true $\theta_0$ and the requirement (R2) as the distribution estimator $H_n(\theta)$ contains right amount of information for $\theta_0$.

The U[0, 1] requirement in (R2) allows us to extract confidence intervals from a CD function easily: $(H_n^{-1}(\alpha_1), H_n^{-1}(1 - \alpha_2))$ is a level $100(1 - \alpha_1 - \alpha_2)\%$ confidence interval for the parameter $\theta_0$, for any

$\alpha_1 > 0$, $\alpha_2 > 0$, and $\alpha_1 + \alpha_2 < 1$. Here, $H_n^{-1}(\beta)$ is the $100\beta\%$ quantile of $H_n(\theta)$ or it solves for $\theta$ in equation $H_n(\theta) = \beta$. CD is a useful device for constructing all types of frequentist statistical inferences, and discussions on extracting information from a CD function to make inferences can be found in Singh, Xie, and Strawderman (2007).

As demonstrated in Singh, Xie, and Strawderman (2005, 2007), the CD concept encompasses a wide range of examples, including most examples in the classical development of Fisher's fiducial distributions, bootstrap distributions, $p$-value functions, standardized likelihood functions, and certain Bayesian prior and posterior distributions. This unification brings together many concepts for statistical inference, and it has both theoretical and practical importance. The following examples of CDs are relevant to our exposition in this article.

*Example A.1* (An aCD for log-odds ratio—a parametric example). Suppose in a binomial clinical trial there are $X_1$ successes and $n_1 - X_1$ failures among $n_1$ patients in the treatment group, and $X_0$ successes and $n_0 - X_0$ failures among $n_0$ patients in the control group. The parameter of interest is the log odds ratio of the treatment $\theta$. The observed log odds ratio is $\hat{\theta} = \log(\{X_1/(n_1 - X_1)\}/\{X_0/(n_0 - X_0)\})$. A well-known asymptotic result (see, e.g., Lehmann 1998, p. 331) suggests that $(\hat{\theta} - \theta)/s \to \mathrm{N}(0, 1)$, as $n_0 \to \infty$ and $n_1 \to \infty$, where $s = \{1/X_0 + 1/(n_0 - X_0) + 1/X_1 + 1/(n_1 - X_1)\}^{1/2}$. By Definition A.1, $H(\theta) = \Phi((\theta - \hat{\theta})/s)$ is an aCD for $\theta$. In another words, the log odds ratio $\theta$ can be estimated by the distribution $\mathrm{N}(\hat{\theta}, s^2)$.

*Example A.2* (Informative prior distribution as a CD). Suppose $\pi(\theta) \sim \mathrm{N}(\mu_0, \sigma_0^2)$ is an informative prior of a parameter $\theta$ of interest. Assume this informative prior is formed on the basis of extensive prior information of $\theta$ from past results of the same or similar experiments. Suppose $Y_0$ is a normally distributed summary statistic from these past experiments, with a realization $Y_0 = \mu_0$ and an observed variance $\mathrm{var}(Y_0) = \sigma_0^2$, respectively. If we denote by $\mathcal{X}$ the sample space of the past experiments and by $\Theta$ the parameter space of $\theta$, we can show by the CD definition that $H_0(\theta) = \Phi((\theta - Y_0)/\sigma_0)$ is a CD function on $\mathcal{X} \times \Theta$. Thus, we consider $H_0(\theta) = \Phi((\theta - Y_0)/\sigma_0) = \Phi((\theta - \mu_0)/\sigma_0)$ as a distribution estimate (CD) from the past experiments. That is, the prior experiments produced $\mathrm{N}(\mu_0, \sigma_0^2)$ as a distribution estimate of $\theta$.

### A.2 Highlights of Recent Developments on CDs

Although the CD concept, especially under the domain of fiducial inference (Fisher 1930; Neyman 1941; Efron 1993; Lehmann 1993), has a long history, it has not received much attention until the recent surge of renewed attention. The recent developments have highlighted CD's promising utility as an effective and powerful tool in statistical inferences. Here are some highlights:

- The renewed interest in CDs starts with Efron (1998), who suggested that bootstrap distributions are "distribution estimators" and CDs. He predicted that "something like fiducial inference" may "become a big hit in the 21st century."
- A new CD concept, under a purely frequentist inference framework, is defined by Schweder and Hjort (2002) and Singh, Xie, and Strawderman (2005). This concept can serve as a unifying framework for many statistical concepts, from regular parametric cases (including most examples in the classical development of Fisher's fiducial distributions) to bootstrap distributions, $p$-value functions, normalized likelihood functions, and, in some cases, Bayesian priors and Bayesian posteriors, and so on. This unifying framework allows us to apply inferences developed in CDs to a broad range of applications.
- Schweder and Hjort (2002) and Singh, Xie, and Strawderman (2005, 2007) explored the connections between likelihood inference and CD-based inference. They also answered several fundamental questions related to the theoretical development of CD inference, including optimality, point estimation, and hypothesis testing, as well as other issues related to decision theory. In addition, Lawless and Fredette (2005) developed a concept of predictive distributions, which can be viewed as a part of CD inference.
- The new developments of CDs also emphasized their applications in modern applied statistics. For examples, Efron (1993) suggested the use of a CD density function to derive an approximate likelihood function. Schweder (2003) developed a CD approach to obtain MLEs of abundance from repeated photographic surveys of a closed stratified population of bowhead whales off Alaska. Tian et al. (2010) used a multivariate CD concept to obtain optimal confidence regions for a vector of constrained parameters. They showed in an analysis of a survival dataset that the volume of the resulting 95% confidence region is only one-thirty-fourth of that of the conventional confidence region. Their confidence region also has better frequency coverage than the corresponding Bayesian credible region. Xie et al. (2009) proposed a frquentist (Bayes compromise) approach to combine experts' opinions with clinical trial data, which is difficult to do in regular frequentist inference without utilizing the CD concept. They demonstrated that the CD-based approach can overcome some inherent drawbacks in the conventional Bayesian approaches for analysis of binomial clinical trials.

## APPENDIX B: PROOFS

### Proof of Lemma 1

Denote $k$ independent U[0, 1] random variables by $U_i$, $i = 1, 2, \ldots, k$. Also let $\tilde{U}_i$, $i = 1, 2, \ldots, k$, be $k$ independent random variables between 0 and 1 such that $|P(\tilde{U}_i \le t) - t| \le \epsilon_i$ for some small positive number $\epsilon_i$ and any $0 < t < 1$. We have, for the first element,

$$\begin{aligned} &P\{G_c(g_c(\tilde{U}_1, \tilde{U}_2, \ldots, \tilde{U}_k)) \le t\} \\ &\quad = E[P\{\tilde{U}_1 \le h_1^{-1}(G_c^{-1}(t)) | \tilde{U}_2, \ldots, \tilde{U}_k\}] \\ &\quad = E[h_1^{-1}(G_c^{-1}(t)) \mathbf{1}_{\{0 \le h_1^{-1}(G_c^{-1}(t)) \le 1\}}] + a_1 \\ &\quad = P\{G_c(g_c(U_1, \tilde{U}_2, \ldots, \tilde{U}_k)) \le t\} + a_1, \end{aligned}$$

where $|a_1| \le \epsilon_1$, $\mathbf{1}_{\{\cdot\}}$ is the indicator function, and $h_1(s) = g_c(s, \tilde{U}_2, \ldots, \tilde{U}_k)$ is the monotonic function of $s$ for a set of given (fixed) $\tilde{U}_2, \tilde{U}_3, \ldots, \tilde{U}_k$. Repeating the same derivation for the 2nd, 3rd, …, $k$th element, we have

$$\begin{aligned} &P\{G_c(g_c(\tilde{U}_1, \tilde{U}_2, \ldots, \tilde{U}_k)) \le t\} \\ &\qquad = P\{G_c(g_c(U_1, U_2, \ldots, U_k)) \le t\} + \sum_{i=1}^{k} a_i, \end{aligned}$$

where $|a_i| \le \epsilon_i$. Replacing $\tilde{U}_i$ by $\tilde{H}_i(\theta_0)$ and $U_i$ by $H_i(\theta_0)$ for $i = 1, 2, \ldots, k$, in the above equation leads to the lemma.

### Proof of Equivalence of Conditions (A) and (A′)

To prove Condition (A) implies Condition (A′), we first note that Condition (A) implies that,

$$|H_i^{-1}(t_2) - H_i^{-1}(t_1)| \to 0 \quad \text{for any fixed } t_1 \text{ and } t_2,\ 0 \le t_1 \le t_2 \le 1, \text{ in probability.} \tag{B.1}$$

This can be proved by taking $\gamma = t_1$ in Condition (A) in the case when $t_2 \le 1 - t_1$ and $\gamma = 1 - t_2$ in the case when $t_2 \ge 1 - t_1$, and noting that $H_i^{-1}(\cdot)$ is a monotonically nondecreasing function.

For any $\epsilon > 0$, taking $t_2 = 1 - \epsilon$, we have $P(H_i(\theta_i^{(0)}) \le t_2) = 1 - \epsilon$. Also, for any $\beta > 0$, taking $t_1 = \beta$ and by (B.1), there exists a large enough $N_0 > 0$ such that, when $n > N_0$, $P(\{H_i(\theta_i^{(0)} - \delta) \ge t_1\} \cap \{H_i(\theta_i^{(0)}) \le t_2\}) \le P(|H_i^{-1}(t_2) - H_i^{-1}(t_1)| > \delta) \le \epsilon$. Since

$P(\{H_i(\theta_i^{(0)} - \delta) \ge t_1\} \cap \{H_i(\theta_i^{(0)}) \le t_2\}) \ge P(H_i(\theta_i^{(0)} - \delta) \ge t_1) + P(H_i(\theta_i^{(0)}) \le t_2) - 1$, it follows that, when $n \ge N_0$,

$$P\big(H_i\big(\theta_i^{(0)} - \delta\big) \ge \beta\big) \le 2\epsilon \quad \text{for any } \epsilon > 0 \text{ and } \beta > 0.$$

Thus, $H_i(\theta_i^{(0)} - \delta) \to 0$, as $n \to \infty$, in probability. Similarly, we prove $1 - H_i(\theta_i^{(0)} + \delta) \to 0$ in probability.

Now we prove Condition (A′) implies Condition (A). For any $\epsilon > 0$ and $0 < \gamma < \frac{1}{2}$, we have, by Condition (A′), $P(\{H_i(\theta_i^{(0)} - \delta) < \gamma\} \cap \{H_i(\theta_i^{(0)} + \delta) > 1 - \gamma\}) \to 1$, as $n \to 0$. It follows that

$$P\big(|H_i^{-1}(1-\gamma) - H_i^{-1}(\gamma)| < 2\delta\big) \to 1,$$

thus Condition (A).

## Proof of Lemma 2

The proof in the case when $F_0$ has bounded support is easy because the function $F_0^{-1}(\cdot)$ is bounded. We consider here the case when $F_0$ has unbounded support. Let $a_\epsilon$ and $b_\epsilon$, for an $\epsilon > 0$, be such that

$$\min_{i \in \mathcal{I}_{1,0}} H_i(a_\epsilon) = \epsilon \qquad \text{and} \qquad \max_{i \in \mathcal{I}_{1,0}} H_i(b_\epsilon) = 1 - \epsilon.$$

Thus, if $a_\epsilon < b_\epsilon$, then $\epsilon \le H_i(\theta) \le 1 - \epsilon$ for all $i \in \mathcal{I}_{1,0}$ and $\theta$ within $[a_\epsilon, b_\epsilon]$. Let us note the bounds

$$P\big(a_\epsilon \ge \theta_1^{(0)}\big) = P\big(H_i\big(\theta_1^{(0)}\big) \le \epsilon \text{ for some } i \in \mathcal{I}_{1,0}\big) \le \|\mathcal{I}_{1,0}\|\epsilon,$$

where $\|\mathcal{I}_{1,0}\|$ is the size of $\mathcal{I}_{1,0}$. Similarly $P(b_\epsilon \le \theta_1^{(0)}) \le \|\mathcal{I}_{1,0}\|\epsilon$. We express, for a $\gamma > 0$,

$$P\Big(\sup_\theta |H_1^{(c)}(\theta) - H_{1,0}^{(c)}(\theta)| > \gamma\Big) \le I + II + III + O(\epsilon),$$

where

$$I = P\Big(\Big\{\sup_{a_\epsilon \le \theta \le b_\epsilon} |H_1^{(c)}(\theta) - H_{1,0}^{(c)}(\theta)| > \gamma\Big\} \cap \{a_\epsilon \le b_\epsilon\}\Big),$$

$$II = P\Big(\sup_{\theta \le a_\epsilon} \max\big\{H_1^{(c)}(\theta), H_{1,0}^{(c)}(\theta)\big\} > \gamma\Big),$$

and

$$III = P\Big(\sup_{\theta \le b_\epsilon} \max\big\{1 - H_1^{(c)}(\theta), 1 - H_{1,0}^{(c)}(\theta)\big\} > \gamma\Big).$$

The argument for the bounded support case applies to $I$; and hence one has

$$I = o(1) + O(\epsilon),$$

where $o(1)$ refers to limit as sample size $n \to \infty$.

We obtain suitable bounds on $II$ and $III$ as follows. Denote by $\mathcal{J}_{1,0} = \{j | H_j(a_\epsilon) = \epsilon, j \in \mathcal{I}_{1,0}\} \subset \mathcal{I}_{1,0}$.

$$\begin{aligned}
&P\Big(\sup_{\theta \le a_\epsilon} H_1^{(c)}(\theta) > \gamma\Big) \\
&= P\Big(\Big\{\sup_{\theta \le a_\epsilon} H_1^{(c)}(\theta) > \gamma\Big\} \\
&\quad \cap \big\{H_i\big(\theta_1^{(0)}\big) \le 1 - \epsilon' \text{ for all } i \in \mathcal{I}_{1,0} - \mathcal{J}_{1,0}\big\} \cap \big\{a_\epsilon \le \theta_1^{(0)}\big\}\Big) \\
&\quad + O(\epsilon) \\
&\le P\Bigg(\Bigg\{G_c\Bigg(\sum_{i \in \mathcal{J}_{1,0}} w_i F_0^{-1}(\epsilon) + \sum_{i \in \mathcal{I}_{1,0} - \mathcal{J}_{1,0}} w_i F_0^{-1}(1-\epsilon) \\
&\quad + \sum_{i \notin \mathcal{I}_{1,0}} w_i F_0^{-1}\big(H_i\big(\theta_1^{(0)}\big)\big)\Bigg) > \gamma\Bigg\} \\
&\quad \cap \big\{H_i\big(\theta_1^{(0)}\big) \le 1 - \epsilon' \text{ for all } i \in \mathcal{I}_{1,0} - \mathcal{J}_{1,0}\big\} \cap \big\{a_\epsilon \le \theta_1^{(0)}\big\}\Bigg) \\
&\quad + O(\epsilon) + O(\epsilon') \\
&= o(1) + O(\epsilon) + O(\epsilon')
\end{aligned}$$

as $\epsilon \to 0$, for a fixed positive $\epsilon'$. Similar arguments are repeated for $H_{1,0}^{(c)}(\theta)$ over $\theta \le a_\epsilon$ and both $1 - H_1^{(c)}(\theta)$ and $1 - H_{1,0}^{(c)}(\theta)$ over $\theta \ge b_\epsilon$. Since $\epsilon' > 0$ is arbitrary, it follows that all $I$, $II$, $III$ tend to zero. This concludes the proof of (i).

To prove (ii), we first note that Conditions (A) and (A′) are equivalent. For $i \in \mathcal{I}_{1,0}$, $H_i(\theta_i^{(0)} - \delta) \to 0$ and $H_i(\theta_i^{(0)} + \delta) \to 1$, in probability. If follows that

$$H_{1,0}^{(c)}\big(\theta_1^{(0)} - \delta\big) \to 0 \qquad \text{and} \qquad H_{1,0}^{(c)}\big(\theta_1^{(0)} + \delta\big) \to 1,$$

in probability. Thus, by theorem 3.1 of Singh, Xie, and Strawderman (2007), $\tilde{\theta}_{1,0}^{(c)} = H_{1,0}^{(c)\,-1}(\frac{1}{2})$ is a consistent estimator of $\theta$. From (i), the result of consistency in (ii) follows.

## Proof of Theorem 1

Since $\hat{\theta}^{(c)}$ is the median, the claim of (i) follows immediately. In the case $\mathcal{I}_0 \ne \emptyset$, from Lemma 2, the claims of (ii) and (iii) hold for all $i \in \mathcal{I}_0$. Therefore they hold for the median as well. In the case when $\mathcal{I}_0 = \emptyset$, which happens when $k = 2m$ is even and $\theta_{(m)}^{(0)} \ne \theta_{(m+1)}^{(0)}$, we have $\theta_0 = (\theta_{(m)}^{(0)} + \theta_{(m+1)}^{(0)})/2$ and $\hat{\theta}^{(c)} = (\hat{\theta}_{(m)}^{(c)} + \hat{\theta}_{(m+1)}^{(c)})/2$. By Lemma 2, we have $\hat{\theta}_{(m)}^{(c)} \to \theta_{(m)}^{(0)}$ and $\hat{\theta}_{(m+1)}^{(c)} \to \theta_{(m+1)}^{(0)}$, the claim of (ii) follows.

## Proof of Theorem 2

Let $n$ denote a generic sample size which is of the order $1/s_i^2$, assuming that all $1/s_i^2$ are of the same order. Without loss of generality, assume that $\theta_1^{(0)} = \theta_0$. Let us express $\hat{\theta}_1^{(c)} - \tilde{\theta}_{1,0}^{(c)}$ as $a/b - c/d$ where

$$a = \sum_1^k \frac{w_i^{(a)}}{s_i^2}(Y_i - \theta_0), \qquad b = \sum_1^k \frac{w_i^{(a)}}{s_i^2},$$

$$c = \sum_{i \in \mathcal{I}_0} \frac{1}{s_i^2}(Y_i - \theta_0), \qquad d = \sum_{i \in \mathcal{I}_0} \frac{1}{s_i^2}.$$

Note that $\mathcal{I}_{1,0} = \mathcal{I}_0$ and $\tilde{\theta}_{1,0}^{(c)} = \tilde{\theta}_0^{(c)}$, and $Y_i$'s have been centered so that $\sqrt{n}|Y_i - \theta_0|$ are bounded in probability. We write now

$$\frac{a}{b} - \frac{c}{d} = a\left(\frac{1}{b} - \frac{1}{d}\right) + \frac{a-c}{d}.$$

The weights corresponding to $i$ outside of $\mathcal{I}_0$ tend to zero at a rate $n^{-k}$, for any $k$ and for those within $\mathcal{I}_0$, $|w_i^{(a)} - 1| = O(n^{-1/2}(\log n)^2)$. Now it takes elementary algebra to conclude that

$$\sqrt{n}a\left(\frac{1}{b} - \frac{1}{d}\right) \to 0 \qquad \text{and} \qquad \sqrt{n}\frac{a-c}{d} \to 0,$$

in probability. In the same way we prove that $\sqrt{n}|\hat{\theta}_j^{(c)} - \tilde{\theta}^{(c)}| \to 0$ for all $j \in \mathcal{I}_0$. So the theorem holds.

## Proof of Theorem 3

Consider the M-estimating equation (4.10). The conclusion (i) follows from the standard argument of an M-estimating equation (see, e.g., Huber 1964).

This proof for (ii) is similar to that of the standard M-estimation as well, except that we need to incorporate the given weight $w_i$ in our case. Note that $H_i(t)$ is bounded between 0 and 1, and maximum contribution of a study to the equation is either $w_i/2$ or $-w_i/2$. In order to

break down the estimating equation so that the solution of the estimating equation approaches infinity, the sum of $w_i$'s over outlying studies should dominate the sum of $w_i$'s over the good studies. Consideration of the worst possible scenario leads to the break down point as stated in (ii).

### Proof of Theorem 4

We only prove the result under random effects model (3.7). The proof under fixed effects model (3.5) is similar. Taylor expansion of the M-estimating equation $\sum_{i=1}^{k}(1/v_i)\{\Phi((\theta - Y_i)/v_i) - \frac{1}{2}\} = 0$ around $Z_i = (\theta_0 - Y_i)/v_i$ yields

$$\hat{\theta}^{(c)} - \theta_0 = \frac{\sum_{i=1}^{k}\{\Phi(Z_i) - 1/2\}/v_i}{\sum_{i=1}^{k}\phi(Z_i)/v_i^2} + o_p\big(n^{-1/2}\big)$$

$$= 2\sqrt{\pi}\,\frac{\sum_{i=1}^{k}(U_i - 1/2)/v_i}{\sum_{i=1}^{k}1/v_i^2} + o_p\big(n^{-1/2}\big),$$

where $U_i$ are independent U[0, 1] random variables and $\phi(\cdot)$ is the density function of the standard normal distribution. The last equation holds because $Z_i$ are independent standard normal random variables and $\mathrm{E}\phi(Z_i) = 1/(2\sqrt{\pi})$. From the expression, we can prove that asymptotically $\mathrm{var}(\hat{\theta}^{(c)}) = \pi/3$. So, the combined CD $H^{(c)}(\theta)$ is asymptotically equivalent to a normal aCD $\Phi((3/\pi)\{\theta - \hat{\theta}^{(c)}\})$. The efficiency claim thus holds.

*[Received December 2009. Revised October 2010.]*

## REFERENCES


Dempster, A. P. (2008), "The Dempster–Shafer Calculus for Statisticians," *International Journal of Approximate Reasoning*, 48, 365–377. [321]

Efron, B. (1993), "Bayes and Likelihood Calculations From Confidence Intervals," *Biometrika*, 80, 3–26. [320,322,325,329,331]

——— (1996), "Empirical Bayes Methods for Combining Likelihoods," *Journal of the American Statistical Association*, 91, 538–550. [321,328,329]

——— (1998), "R. A. Fisher in the 21st Century," *Statistical Science*, 13, 95–122. [320,324,331]

Fisher, R. A. (1930), "Inverse Probability," *Proceedings of the Cambridge Philosophical Society*, 26, 528–535. [331]

——— (1932), *Statistical Methods for Research Workers* (4th ed.), Edinburgh, Scotland: Oliver & Boyd. [322]

——— (1973), *Statistical Methods and Scientific Inference* (3rd ed.), New York: Hafner Press. [320]

Fraser, D. A. S. (1991), "Statistical Inference: Likelihood to Significance," *Journal of the American Statistical Association*, 86, 258–265. [320,322,323]

Hannig, J. (2009), "On Generalized Fiducial Inference," *Statistica Sinica*, 19, 491–544. [321]

Hedges, L. V., and Olkin, I. (1985), *Statistical Methods for Meta-Analysis*, New York: Academic Press. [320]

Hu, F., and Zidek, J. V. (2002), "The Weighted Likelihood," *Canadian Journal of Statistics*, 30, 347–371. [327]

Huber, P. J. (1964), "Robust Estimation of a Location Parameter," *Annals of Mathematical Statistics*, 35, 73–101. [326,332]

Lawless, F., and Fredette, M. (2005), "Frequentist Prediction Intervals and Predictive Distributions," *Biometrika*, 92 (3), 529–542. [320,331]

Lehmann, E. L. (1993), "The Fisher, Neyman–Pearson Theories of Testing Hypotheses: One Theory or Two?" *Journal of the American Statistical Association*, 88, 1242–1249. [331]

——— (1998), *Elements of Large-Sample Theory*, New York: Springer-Verlag. [331]

Littell, R. C., and Folks, J. L. (1973), "Asymptotic Optimality of Fisher's Method of Combining Independent Tests. II," *Journal of the American Statistical Association*, 68, 193–194. [322]

Marden, J. I. (1991), "Sensitive and Sturdy $p$-Values," *The Annals of Statistics*, 19, 918–934. [322,323]

Martin, R., Zhang, J., and Liu, C. (2010), "Dempster–Shafer Theory and Statistical Inference With Weak Beliefs," *Statistical Science*, 25, 72–87. [321]

Neyman, J. (1941), "Fiducial Argument and the Theory of Confidence Intervals," *Biometrika*, 32, 128–150. [331]

Normand S.-L. (1999), "Meta-Analysis: Formulating, Evaluating, Combining, and Reporting," *Statistics in Medicine*, 18, 321–359. [321-324,328,330]

Parzen, E. (2005), "All Statistical Methods, Parameter Confidence Quantiles," in *Noether Award Lecture at the Joint Statistical Meeting*, Minneapolis, MN, U.S.A., August 7–11, 2005. [320]

Preiss, R. W., Gayle, B. M., Burrell, N., Allen, M., and Bryant, J. (2006), *Mass Media Effects Research: Advances Through Meta-Analysis*, New York: Springer-Verlag. [320]

Schulze, R. (2004), *Meta-Analysis: A Comparison of Approaches*, Cambridge, MA: Hogrefe & Huber. [320]

Schweder, T. (2003), "Abundance Estimation From Multiple Photo Surveys: Confidence Distributions and Reduced Likelihoods for Bowhead Whales off Alaska," *Biometrics*, 59, 974–983. [320,331]

Schweder, T., and Hjort, N. L. (2002), "Confidence and Likelihood," *Scandinavian Journal of Statistics*, 29, 309–332. [320,330,331]

——— (2003), "Frequentist Analogues of Priors and Posteriors," in *Econometrics and the Philosophy of Economics*, ed. B. P. Stigum, Princeton, NJ: Princeton University Press, pp. 285–317. [320]

Singh, K., Xie, M., and Strawderman, W. E. (2005), "Combining Information From Independent Sources Through Confidence Distributions," *The Annals of Statistics*, 33, 159–183. [320-322,324-326,330,331]

——— (2007), "Confidence Distribution (CD)—Distribution Estimator of a Parameter," in *Complex Datasets and Inverse Problems. IMS Lecture Notes—Monograph Series*, Vol. 54, eds. R. Liu et al., Beachwood, OH: Institute of Mathematical Statistics, pp. 132–150. [320,322,323,325,331,332]

Stangl, D. K., and Berry, D. A. (2000), *Meta-Analysis in Medicine and Health Policy*, New York: Springer-Verlag. [320]

Stouffer, S. A., Suchman, E. A., DeVinney, L. C., Star, S. A., and Williams, R. M., Jr. (1949), *Adjustment During Army Life*, Princeton, NJ: Princeton University Press. [322]

Sutton, A. J., and Higgins, J. P. T. (2008), "Recent Developments in Meta-Analysis," *Statistics in Medicine*, 27, 625–650. [320,323]

Sweeting, M. J., Sutton, A. J., and Lambert, P. C. (2004), "What to Add to Nothing? Use and Avoidance of Continuity Corrections in Meta-Analysis of Sparse Data," *Statistics in Medicine*, 23, 1351–1375. [329]

Tian, L., Wang, R., Cai, T., and Wei, L. J. (2010), "The Highest Confidence Density Region and Its Usage for Joint Inferences About Constrained Parameters," *Biometrics*, to appear. DOI: *10.1111/j.1541-0420.2010.01486.x*. [320,331]

Wang, X., and Zidek, J. V. (2005), "Selecting Likelihood Weights by Cross-Validation," *The Annals Statistics*, 33, 463–501. [327]

Whitehead, A. (2002), *Meta-Analysis of Controlled Clinical Trials*, London: Wiley. [320]

Xie, M., Liu, R. Y., Damaraju, C. V., and Olson, W. H. (2009), "Incorporating Expert Opinions With Information From Binomial Clinical Trials," technical report, Rutgers University, Dept. of Statistics. [320,331]

Yang, G., and Xie, M. (2010), "*gmeta*: Meta-Analysis via a Unified Framework Under Confidence Distribution," R package, available at *http://stat.rutgers.edu/home/mxie/packages/gmetaRpackage/*. [330]