CONFIDENCE INTERVALS FOR POPULATION RANKS IN THE PRESENCE OF TIES AND NEAR TIES

Minge Xie, Kesar Singh and Cun-Hui Zhang

Abstract

Frequentist confidence intervals for population ranks and their statistical justifications are not well established, even though there is a great need for such procedures in practice. How do we assign confidence bounds for the ranks of health care facilities, schools and financial institutions based on data that do not clearly separate the performance of different entities apart? The commonly used bootstrap-based frequentist confidence intervals and Bayesian intervals for population ranks may not achieve the intended coverage probability in the frequentist sense, especially in the presence of unknown ties or "near ties" among the populations to be ranked. Given random samples from k populations, we propose confidence bounds for population ranking parameters and develop rigorous frequentist theory and non-standard bootstrap inference for population ranks which allow ties and near ties. In the process, a notion of modified population rank is introduced which appears quite suitable for dealing with the population ranking problem. The proposed methodology and theoretical results are illustrated through simulations and a real dataset from a health research study involving 79 VHA facilities. The results are extended to general risk adjustment models.

Keywords and Phrases: Large sample theory, Nonstandard Bootstrap inference, Population rank, Rank inference, Slow convergence rate, Ties and near ties

1 Introduction

Performance evaluations of institutions are common and have a long history in many areas. Aitkin and Longford (1986) and Goldstein and Spiegelhalter (1996) discussed the importance of introducing accountability measurements to public sector activities and spelled out the statistical issues involved in making quality comparisons among different institutions in the areas of health care and education. These publications have led to many applications and developments in health services (e.g., Christiansen and Morris, 1997; Normand et al. 1997; Landrum et al. 2003; Thompson et al. 2005; Austin and Tu 2006), education (e.g., Lockwood et al. 2002; McCaffrey et al. 2004; Rubin et al. 2004; Noell and Burns 2006), and many other fields. Goldstein and Spiegelhalter (1996) pointed out that performance evaluation "inevitably leads to institutional ranking" and repeatedly emphasized the need for "interval estimation" to display uncertainty associated with such rankings. According to Goldstein and Spiegelhalter (1996), there are generally two computational approaches for deriving confidence intervals for population ranks: Bayesian computational method and frequentist bootstrap sampling method.

An earlier paper by Laird and Louis (1989) and the follow-up papers by Shen and Louis (1998) and Lin et al. (2006) proposed several Bayesian methods and studied from theoretical aspects Bayesian inference for ranks in hierarchical models. Theoretical discussion of frequentist inference for rank parameters, however, is rare in the literature. A main reason for the lack of frequentist studies of the problem is that the population rank is a "difficult" parameter for which the standard statistical theory does not apply, as pointed out by Snijders (1996). We study in this paper, within the frequentist domain, the statistical inference for population ranks, especially confidence intervals in the presence of unknown ties or near ties of population ranks.

Suppose we are interested in ranking k populations (or institutions) $\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_k$ through a specific characteristic described by a set of unknown parameter values $\theta_1, \ldots, \theta_k$. The conventional frequentist definition of the population rank for $\mathcal{P}_i$ is then

$$R_i = 1 + \sum_{j \neq i} G(\theta_j - \theta_i) = 1 + \sum_{j \neq i} \mathbf{1}_{(\theta_j < \theta_i)} + \frac{1}{2} \sum_{j \neq i} \mathbf{1}_{(\theta_j = \theta_i)}, \quad (1)$$

where $G(t) = \mathbf{1}_{(t<0)} + \frac{1}{2}\mathbf{1}_{(t=0)}$, and $\mathbf{1}_{(A)}$ is the 0-1 indicator function of A.

Suppose consistent estimates $\hat{\theta}_{in}$ of the unknown parameters θ_i are available, with $\hat{\theta}_{in} = \theta_i + o_P(1)$, where n is a known constant representing a generic "sample size" or the order of the reciprocal of the variances of $\hat{\theta}_{in}$. For instance, n is just the number of observations taken from each population in case of equal size. Or, in the case of unequal sample sizes, n_i for population i, we may take $n = \sum_{i=1}^{k} n_i$ as the total sample size for fixed k. Alternatively, we may take $n = n_1$ if the rank of $\mathcal{P}_1$ is of special interest, provided $\max_{i\le k} n_i/n_1 = O(1)$. A commonly used plug-in estimator of the population rank (1) is defined as

$$\hat{R}_{in} = 1 + \sum_{j\neq i} G(\hat{\theta}_{jn} - \hat{\theta}_{in}). \tag{2}$$

When k is fixed and $\theta_1, \theta_2, \ldots, \theta_k$ are all fixed and distinct, it can be proved that the consistency of $\hat{\theta}_{in}$ implies the consistency of the plug-in estimator in the sense of

$$P(\hat{R}_{in} = R_i) \to 1 \quad \text{as } n \to \infty, \tag{3}$$

so that $\hat{R}_{in}$ could serve as a "singleton confidence set" for the population rank R_i with asymptotically perfect coverage. Although this theory is simple, it does not agree well with most practical settings since it requires that the parameters θ_i be well separated. This mismatch between (3) and reality is clearly exhibited in our numerical experiments reported in Sections 4 and 5. Indeed, one of the main reasons that cause controversies in ranking practices is that the performances of some institutions are not easily distinguishable, in addition to risk adjustment and measurement error issues.

A more relevant problem, which we focus in this paper, is to construct confidence intervals of population ranks which (approximately) achieve the claimed coverage probability when the performances of the institutions to be ranked are not always well separated. The population parameters θ_i may have ties or near ties, under which the statement (3) breaks down completely. Here a near tie between parameters θ_i and θ_j means that their difference is not of greater order than the standard error of the estimate of the difference:

$$\left(\hat{\theta}_{in} - \hat{\theta}_{jn}\right) - \left(\theta_i - \theta_j\right) \neq o_P(|\theta_i - \theta_j|). \tag{4}$$

That is, we can not distinguish θ_i and θ_j with the available data. In the special case when both $\hat{\theta}_{in}$ and $\hat{\theta}_{jn}$ are root-n consistent, (4) can be written in a more straightforward expression $|\theta_i - \theta_j| = O(n^{-\frac{1}{2}})$. Of course, near ties include ties $\theta_i = \theta_j$ and the label pairs (i, j) of ties or near ties are assumed completely unknown.

Table 1 displays a data set from a health research study on blood glucose (A1c) control of diabetes patients in 79 Veteran Administration Facilities (hospitals) in the United States (Pogach et al., 2005). Figure 1 plots log odds of "poor A1c control" rates against the facility ids (sorted by the facility sample ranks $\hat{R}_{in}$). As in Pogach et al. (2005), the "poor Alc control" for a diabetes patient is defined as an Alc level over 9.5 or an Alc test not performed. From Figure 1, it is clear that the log odds or the "poor A1c control" rates are clumped together. Even though the sample sizes (n_i reported in Table 1) are in the thousands, most of the reported rates are not distinguishable from their neighboring points by any statistical test. In this example, the assumption under which (3) holds is no longer valid, and (3) is not expected to yield reliable confidence intervals for the population ranks.

Goldstein and Spiegelhalter (1996) suggested bootstrap as a method of constructing confidence intervals for population ranks. Commonly used bootstrap confidence intervals rely on the centered bootstrap percentile approach (where the intervals are obtained by approximating the distribution of $\hat{R}_i - R_i$ by the distribution of $\hat{R}_i^* - \hat{R}_i$) and the regular bootstrap percentile method (where the bounds are the quantiles of the bootstrap distribution); see, Efron and Tibshirani (1994). Here, $\hat{R}_i^*$ is the sample rank computed from a bootstrap sample. These two approaches are equivelant, if the bootstrap distribution of the rank is symmetric. But they need not be the same otherwise. The intervals from the regular bootstrap percentile method turn out to be also asymptotically equivalent to Bayesian credible intervals. It is well known that Bayesian credible intervals do not guarantee the frequency coverage of confidence intervals; see, e.g., Zhang and Woodroofe (2002) and Marchand and Strawderman (2006) for studies of frequency coverage issues of Bayesian credible intervals. Although it is not completely clear theoretically whether the conventional bootstrap methods or a Bayesian method yield confidence intervals with intended coverage probability for the population ranks, we have evidence in Section 2 that the regular bootstrap theory breaks down. Our simulation study reported in Section 5 also strongly suggests that the intervals obtained by the aforementioned conventional bootstrap methods have poor frequentist coverage probabilities.

In the presence of ties or near ties, the sample rank $\hat{R}_{in}$ in (2) is not a consistent estimator of the population rank R_i in (1). To deal with this tie and near tie problems, we propose a smooth population rank $R_i^{(smooth)}$ and a smooth sample rank $\hat{R}_{in}^{(smooth)}$ with the following properties. For fixed distinct parameters $\theta_1, \ldots, \theta_k$ or in the presence of ties, the proposed $\hat{R}_{in}^{(smooth)}$ consistently estimates the conventional population rank R_i and the estimation error $\hat{R}_{in}^{(smooth)} - R_i$ is asymptotically normal under mild conditions, provided absence of near ties to θ_i. In these cases, the difference between the smooth population rank $R_{in}^{(smooth)}$ and conventional population rank R_i typically diminishes exponentially fast. Based on the asymptotic normality of $\hat{R}_{in}^{(smooth)}$, we propose bootstrap confidence intervals which provide consistent coverage probability for the population ranks R_i and $R_i^{(smooth)}$. In the presence of near ties as defined in (4), the problem is much more complicated. In these cases, the above statements still hold for $R_{in}^{(smooth)}$ (in place of R_i), but they no longer hold for the conventional population rank R_i. A conservative confidence interval for the conventional population rank is then proposed. Although the smooth rank is generally not an integer and thus might not satisfy some application purposes, our numerical results, based on simulated and real datasets, support the above claims and demonstrate the superior performance of the newly proposed procedures compared with the existing ones.

Section 2 below introduces the smooth ranks and statistical procedures and study their properties. Section 3 extends the results to risk adjustment models. Sections 4 illustrates the proposed methodology using the VHA data. Section 5 examines the performance of the proposed procedures using simulations. Section 6 contains some further discussions.

2 Ranks and Bootstrap Inferences

We assume throughout this section that the data provide certain mutually independent root-n consistent estimators $\hat{\theta}_{in}$ of the population parameters θ_i. In particular, we assume

$$\hat{\theta}_{in} = \theta_i + n^{-1/2}\Big(Z_{in} + o_P(1)\Big), \quad Z_{in} \sim N(0, \sigma_i^2), \quad i \leq k, \tag{5}$$

with scale constants σ_i satisfying $0 < \sigma_* \leq \min_{i \leq k} \sigma_i \leq \max_{i \leq k} \sigma_i \leq \sigma^* < \infty$. For simplicity, we assume that k, the total number of populations to be compared, is fixed, although this restriction can be removed under mild further assumptions on the $o_P(1)$ in (5). The population parameters θ_i are allowed to depend on n in the sense that all the statements in this section are uniform in their values.

2.1 Confidence intervals for conventional ranks in the presence of ties

Suppose without loss of generality that we are interested in statistical inference about the rank R_1 of the first population characterized by θ_1. Let

$$\Theta_T = \left\{ j : \theta_j = \theta_1,\ j = 2, 3, \ldots, k \right\}$$

be the set of populations that are tied to the first population. Here, Θ_T is completely unknown, not even whether it is empty or not. We assume throughout this subsection that

$$\kappa_n \equiv \sqrt{n} \min_{j \notin \Theta_T, j > 1} \left| \theta_j - \theta_1 \right| \to \infty. \tag{6}$$

The interpretation of (6) is that the parameters θ_j are either well separated from θ_1 or tied to θ_1; i.e., there is no near tie to the first population.

The conventional population and sample ranks for the first population, R_1 and $\hat{R}_1$ respectively, are defined by (1) and (2) with $i = 1$. It follows from (2) and (5) that

$$\hat{R}_1 = 1 + \sum_{j=2}^{k} \mathbf{1}_{\left(Z_{jn} - Z_{1n} \leq \sqrt{n}(\theta_1 - \theta_j) + o_p(1) \right)}. \tag{7}$$

For $\Theta_T = \emptyset$, it follows form (7) that $P(\hat{R}_1 = R_1) \to 1$ under the conditions (5) and (6). Thus, when there is no tie to the first population, $\hat{R}_1$ is a consistent estimator of R_1 and confidence intervals for R_1 at any level always have asymptotically 100% coverage probability as long as the interval contains the estimator $\hat{R}_1$, regardless of the methods used.

For $\Theta_T \neq \emptyset$, however, the story is totally different. For a population with θ_j tied to θ_1, its contribution to the population rank R_1 is 1/2, but its contribution to the plug-in estimator $\hat{R}_1$ converges to a Bernoulli variable with probability 1/2 under (5), resulting in

$$\hat{R}_1 - R_1 = \sum_{j \in \Theta_T} \left\{ \mathbf{1}_{\left(Z_{jn} - Z_{1n} \leq o_P(1)\right)} - \frac{1}{2} \right\} + o_p(1).$$

Thus, the sample rank $\hat{R}_1$ is not a consistent estimator of the population rank R_1 in the presence of ties. In fact, $\hat{R}_1 - R_1 + ||\Theta_T||/2$ converges in distribution to a binomial distribution, where $||\Theta_T||$ denotes the number of elements in Θ_T. In this case, the regular bootstrap theory breaks down and the conventional frequentist bootstrap confidence interval suggested by Goldstein and Spiegelhalter (1996) is no longer valid.

To overcome the difficulty, we propose a smooth sample rank

$$\hat{R}_1^{(smooth)} = 1 + \sum_{j=2}^{k} G_n(\hat{\theta}_{jn} - \hat{\theta}_{1n}) = k - \sum_{j=2}^{k} F\left(\frac{\hat{\theta}_{jn} - \hat{\theta}_{1n}}{\tau_n}\right) \tag{8}$$

as an estimate of the population rank, where $G_n(t) = 1 - F(t/\tau_n)$ with a continuous CDF $F(\cdot)$, and τ_n are positive constants satisfying

$$\tau_n \to 0 \quad \text{and} \quad \sqrt{n}\tau_n \to +\infty. \tag{9}$$

Our theoretical results only require $F(0) = 1/2$ and the uniform continuous differentiability of F. It can be seen that, for each fixed t, $G_n(t) \to G(t) = \mathbf{1}_{(t<0)} + \frac{1}{2}\mathbf{1}_{(t=0)}$ as $\tau_n \to 0$.

A motivation of our proposal of the smooth estimator $\hat{R}_1^{(smooth)}$ in (8) is its connection to the p-values

$$p_j = 1 - \Phi\left(\frac{\hat{\theta}_{jn} - \hat{\theta}_{1n}}{n^{-1/2}(\sigma_j^2 + \sigma_1^2)^{1/2}}\right)$$

for testings H_0: $\theta_j \leq \theta_1$ against H_1: $\theta_j > \theta_1$. The p-value p_j can be viewed as a confirmation value for the inequality $\theta_j \leq \theta_1$. If σ_j are all identical and τ_n is taken to be $n^{-1/2}\sigma_1\sqrt{2}$, then (8) becomes $1 + \sum_{j=2}^{n} p_j$ as the sum of the p-values. However, the sum of the p-values is not smooth enough for the theoretical justification of confidence procedures for the population rank. This leads to the introduction of the condition $\sqrt{n}\tau_n \to +\infty$ in (9). The proposed estimator is also related to the empirical Bayesian rank derived by Laird and Louis (1989) using a hierarchical model. This connection will be discussed in the next subsection.

The next theorem states that $\hat{R}_1^{(smooth)}$ is a consistent estimator of the population rank R_1 even in the presence of ties, and that it can be utilized to construct confidence intervals for R_1. This is a non-standard asymptotic normality result since the convergence rate $n^{-1/2}/\tau_n$ is slower than the regular rate $n^{-1/2}$.

Theorem 2.1 *Let Θ_T be the set of ties to the parameter θ_1 of the first population. Suppose the distribution function F in (8) is uniformly continuously differentiable with $F(0) = 1/2$. Suppose $\kappa_n \to \infty$ as in (6). Let τ_n be constants satisfying (9) and*

$$1 - F\Big(\kappa_n/(n^{1/2}\tau_n)\Big) + F\Big(-\kappa_n/(n^{1/2}\tau_n)\Big) = \frac{o(1)}{n^{1/2}\tau_n}. \tag{10}$$

(i) Denote by $v_1 = (F'(0))^2(||\Theta_T||^2\sigma_1^2 + \sum_{j\in\Theta_T}\sigma_j^2)$. Then, as $n \to \infty$,

$$\begin{cases} \tau_n\sqrt{n}\, v_1^{-1/2}\Big(\hat{R}_1^{(smooth)} - R_1\Big) \to_D N(0,1), & \textit{if } v_1 \neq 0; \\ \tau_n\sqrt{n}\Big(\hat{R}_1^{(smooth)} - R_1\Big) \to_P 0, & \textit{if } v_1 = 0. \end{cases}$$

Consequently, the estimator (8) is consistent in the sense of $\hat{R}_1^{(smooth)} - R_1 = o_P(1)$.
(ii) Suppose $\hat{\theta}_{jn}^*$ are computed from bootstrap samples such that $n^{1/2}(\hat{\theta}_{jn}^* - \hat{\theta}_{jn}) \sim^d n^{1/2}(\hat{\theta}_{jn} - \theta_j)$ in limit, almost surely. Let $\hat{R}_1^{(smooth)^*} = 1 + \sum_{j=2}^k G_n\big(\hat{\theta}_{jn}^* - \hat{\theta}_{1n}^*\big)$ be the smooth rank of the first population based on the bootstrap estimates $\{\hat{\theta}_{jn}^*, 1 \le j \le k\}$. Then, the bootstrap distribution of $\tau_n\sqrt{n}\big(\hat{R}_1^{(smooth)^*} - \hat{R}_1^{(smooth)}\big)$ consistently estimates the distribution of $\tau_n\sqrt{n}\big(\hat{R}_1^{(smooth)} - R_1\big)$ in the case when $v_1 \neq 0$. In the case when $v_1 = 0$, this bootstrap result still holds, provided that $F = \Phi$ and $\min_{j\neq 1}|\theta_j - \theta_1|/(\tau_n^2 n^{1/2}) \to \infty$.

The proof of Theorem 2.1 can be found in Appendix.

The κ_n defined in (6) can be treated as a characteristic parameter of the populations that can not change for a given data set, but the parameter τ_n is tunable which is similar to the tuning parameter "bandwidth" in (kernel) density estimation or nonparametric regressions. In the case of no near tie in the sense of (6), there always exists a sufficiently wide range of τ_n satisfying (9) and the upper bound requirement (10). Note that, under the moment condition $\int |x|^\alpha dF(x) < \infty$ for any $\alpha > 0$, any $\tau_n \le n^{-1/2}\kappa_n^{\alpha/(1+\alpha)}$ yields

$$1 - F\Big(\kappa_n/(n^{1/2}\tau_n)\Big) + F\Big(-\kappa_n/(n^{1/2}\tau_n)\Big) = o\Big(\frac{n^{1/2}\tau_n}{\kappa_n}\Big)^\alpha = \frac{o(1)}{n^{1/2}\tau_n}.$$

For the choice of $F(x) = \Phi(x)$ in particular, a τ_n such that $\kappa_n \ge n^{1/2}\tau_n\sqrt{2\log(n^{1/2}\tau_n)}$ suffices for (10). Thus, if $\theta_j \notin \Theta_T$ are at a non-shrinking distance from θ_1, a convenient choice for implementation is $\tau_n \propto n^{-\beta/2}$, for some $0 < \beta < 1$. See Section 4 for further discussions and a specific choice of the tuning parameter τ_n.

2.2 Smooth ranks in the presence of near ties

The term "near tie" refers to two populations whose true characteristic parameters differ by a gap of the same order as the estimation error. Thus, under the condition (5),

$$\Theta_N = \left\{ j : |\theta_j - \theta_1| = O(n^{-\frac{1}{2}}),\ j = 2, 3, \ldots, k \right\} \tag{11}$$

is the set of near ties to the first population. Since it is impossible to distinguish ties from near ties based on data under (5), it is convenient to include true ties with $\theta_j = \theta_1$ as near ties: $\Theta_N \supset \Theta_T$. Again, the size and membership of Θ_N are assumed to be completely unknown, not if members of Θ_N are true ties or not, not even if Θ_N is empty or not.

The contribution of a near-tie population with $j \in \Theta_N$ to the rank R_1 of the first population ranges anywhere among 0, 1/2 and 1, depending on the value of the difference $\theta_j - \theta_1$ ranging within the order of measurement error. It is not possible to find a consistent estimator of R_1 under the asymptotic condition (5). Although Theorem 2.1 is no longer valid, the expansion (A.2) in its proof in the Appendix exhibits that the smooth sample rank $\hat{R}_1^{(smooth)}$ of (8) well approximates the smooth population rank

$$R_1^{(smooth)} = 1 + \sum_{j=2}^{k} G_n(\theta_j - \theta_1) = k - \sum_{j=2}^{k} F\Big((\theta_j - \theta_1)/\tau_n\Big) \tag{12}$$

where $G_n(\cdot)$, F and τ_n are as in (8). In the case of no near tie ($\Theta_N - \Theta_T = \emptyset$), we have $R_1^{(smooth)} - R_1 \to 0$. But in the presence of near ties ($\Theta_N - \Theta_T \neq \emptyset$), $R_1^{(smooth)} - R_1 \not\to 0$.

The smooth population rank $R_1^{(smooth)}$ is particularly convenient to deal with, compared with the discontinuous conventional rank R_1. Furthermore, $R_1^{(smooth)}$ has an interesting Bayesian connection under the following hierarchical prior:

$$\{\theta_j\}\Big|\{\mu_j\} \sim \text{ independent } N(\mu_j, \xi_1), \quad \{\mu_j\} \sim \text{ iid } N(\mu, \xi_2).$$

Under this hierarchical prior, it is reasonable to think of $\tilde{R}_1 = 1 + \sum_{j=2}^{k} \mathbf{1}_{(\mu_j < \mu_1)}$ as the desired ranking for the first population, and since $\tilde{R}_1$ is not available

$$R_1^{(smooth)} = E\Big[\tilde{R}_1\Big|\theta_1, \ldots, \theta_n\Big]$$

as the next choice, due to $\mu_j|\theta_j \sim N((\mu\xi_1+\theta_j\xi_2)/(\xi_1+\xi_2), \xi_1\xi_2/(\xi_1+\xi_2))$, provided $\tau_n = \sqrt{\xi_1(1+\xi_1/\xi_2)}$. This is closely related to Laird and Louis (1989), but different from their Bayesian notion, in view of the formula for τ_n.

Theorem 2.2 below states that $\hat{R}_1^{(smooth)}$ is a consistent estimator of the smooth population rank $R_1^{(smooth)}$, and bootstrap inference can be applied to them. Theorem 2.2 certainly covers the case of no near tie ($\Theta_N = \Theta_T$). In fact, Theorem 2.1 is a consequence of Theorem 2.2 and the fact that $R_1 - R_1^{(smooth)} = o(n^{-1/2}/\tau_n)$ in the absence of near ties.

Theorem 2.2 *Let Θ_N be the set of near ties to the parameter θ_1 of the first population as in (11). Suppose the distribution function F in (8) is uniformly continuously differentiable with $F(0) = 1/2$. Let τ_n be constants satisfying (9) and*

$$1 - F\Big(\kappa_n^*/(n^{1/2}\tau_n)\Big) + F\Big(-\kappa_n^*/(n^{1/2}\tau_n)\Big) = \frac{o(1)}{n^{1/2}\tau_n} \tag{13}$$

where $\kappa_n^* \equiv \sqrt{n}\min_{j\notin\Theta_N, j\geq 2}|\theta_j - \theta_1| \to \infty$ as $n \to \infty$.
(i) Denote by $v_{1n}^* = (F'(0))^2(||\Theta_N||^2\sigma_1^2 + \sum_{j\in\Theta_N}\sigma_j^2)$. Then, as $n \to \infty$,

$$\begin{cases} \tau_n\sqrt{n}\,(v_{1n}^*)^{-1/2}\Big(\hat{R}_1^{(smooth)} - R_1^{(smooth)}\Big) \to_D N(0,1), & \text{if } v_{1n}^* \neq 0; \\ \tau_n\sqrt{n}\Big(\hat{R}_1^{(smooth)} - R_1^{(smooth)}\Big) \to_P 0, & \text{if } v_{1n}^* = 0. \end{cases}$$

Consequently, (8) is consistent in the sense of $\hat{R}_1^{(smooth)} - R_1^{(smooth)} = o_P(1)$.
(ii) Let $\hat{\theta}_{in}^*$ and $\hat{R}^{(smooth)*}$ be the bootstrap estimates as in Theorem 2.1 (ii). Then, the bootstrap distribution of $\tau_n\sqrt{n}\big(\hat{R}_1^{(smooth)*} - \hat{R}_1^{(smooth)}\big)$ consistently estimates the distribution of $\tau_n\sqrt{n}\big(\hat{R}_1^{(smooth)} - R_1^{(smooth)}\big)$ in the case when $v_{1n}^* \neq 0$. In the case when $v_{1n}^* = 0$, this bootstrap result still holds, provided that $F = \Phi$ and $\min_{j\neq 1}|\theta_j - \theta_1|/(\tau_n^2 n^{1/2}) \to \infty$.

The proof of Theorem 2.2 is omitted since it is nearly identical to and slightly simpler than the proof of Theorem 2.1.

In Theorem 2.2, the unknown parameters $\theta_j \equiv \theta_{jn}$ are allowed to depend on n. It is required that for each j either $\sup_n \sqrt{n}\big|\theta_{jn} - \theta_{1n}\big| < \infty$ or $\sqrt{n}\big|\theta_{jn} - \theta_{1n}\big| \to \infty$. In this case, $\Theta_N = \{j : \sup_n \sqrt{n}\big|\theta_{jn} - \theta_{1n}\big| < \infty\}$ by (11) and $\kappa_n^* \to \infty$ due to the finiteness of the total number k of populations to be ranked. It is possible to extend the theorem to the case of $k = k(n) \to \infty$ under stronger versions of (5) and (13), provided the existence of $\Theta_N = \Theta_{N,n}$ satisfying $\kappa_n^* \to \infty$ and $\sqrt{n}\max_{j\in\Theta_N} |\theta_{jn} - \theta_{1n}| = O(1)$. Again, the discussion following Theorem 2.1 on the characteristic parameter κ_n and the tuning parameter τ_n still holds, but with κ_n replaced by κ_n^*. If $\theta_j \notin \Theta_N$ are at a non-shrinking distance from θ_1 and for the choice of $F(x) = \Phi(x)$, a convenient choice of τ_n is again $\tau_n \propto n^{-\beta/2}$, for some $0 < \beta < 1$; further discussions and implementation of the method can be found in Section 4.

2.3 Confidence bounds for conventional ranks in the presence of near ties

As mentioned in Subsection 2.2, the conventional population rank R_1 is intrinsically discontinuous in the presence of near ties. Still, the smoothing methodology that we developed can be used to construct confidence bounds for R_1, even in the presence of near ties. The idea is to examine the difference $R_1 - R_1^{(smooth)}$, and obtain a set of upper and lower bounds for the difference. Then, using these bounds and confidence intervals of $R_1^{(smooth)}$, we obtain confidence intervals for R_1.

The difference between R_1 and $R_1^{(smooth)}$ can be expressed as

$$\begin{aligned} R_1 - R_1^{(smooth)} &= \sum_{j\in\Theta_N} \left\{ F\left(\frac{\theta_j - \theta_1}{\tau_n}\right)\mathbf{1}_{(\theta_j<\theta_1)} - \bar{F}\left(\frac{\theta_j - \theta_1}{\tau_n}\right)\mathbf{1}_{(\theta_j>\theta_1)} \right\} + o(1) \\ &= \sum_{j\in\Theta_N} \frac{1}{2}\left\{ 1_{(\theta_j<\theta_1)} - 1_{(\theta_j>\theta_1)} \right\} + o(1), \end{aligned}$$

where $\bar{F}(t) = 1 - F(t)$. It leads to a set of inequalities for the difference of $R_1 - R_1^{(smooth)}$

$$-\frac{1}{2}||\Theta_N|| \le R_1 - R_1^{(smooth)} \le \frac{1}{2}||\Theta_N||, \tag{14}$$

where $||\,\Theta_N||$ is the number of near ties in Θ_N. However, $||\,\Theta_N||$ is not known. To estimate it, we introduce a continuous kernel function $K(x)$. It is required that the kernel $K(x) \ge 0$ for any $x \in (-\infty, +\infty)$, $K(0) = 1$ and, as $|x| \to \infty$, $K(x) \to 0$ at a certain rate. Our choice is $K(t) = 2\,\min\{F(t), \bar{F}(t)\}$, which maintains the same tail convergence rate as the function $F(t)$ in the smooth estimator $\hat{R}_1^{(smooth)}$.

The following theorem states that $\widehat{||\Theta_N||} = \sum_{j=2}^k K\big((\hat{\theta}_j - \hat{\theta}_1)/\tau_n\big)$ is a consistent estimator of $||\Theta_N||$. It also provides confidence intervals for the conventional rank R_1. Here, L_1^* and U_1^* are the lower and upper bounds of a level $100\gamma\%$ confidence interval of the smooth rank $R_1^{(smooth)}$, respectively. Also, the $floor(x)$ function rounds x to the largest integer not greater than x, and the $ceiling(x)$ function rounds x to the smallest integer not less than x.

Theorem 2.3 *Under the conditions of Theorem 2.2, we have*

$$\widehat{||\Theta_N||} = \sum_{j=2}^{k} K\left(\frac{\hat{\theta}_j - \hat{\theta}_1}{\tau_n}\right) = ||\Theta_N|| + o_p(1)$$

and a set of confidence intervals for the conventional rank

$$floor\left(L_1^* + \frac{1}{2}\widehat{||\Theta_N||}\right) \leq R_1 \leq ceiling\left(U_1^* + \frac{1}{2}\widehat{||\Theta_N||}\right).$$

It is guaranteed that, asymptotically, the coverage of this interval is at least $100\gamma\%$.

The proof of Theorem 2.3 can be found in Appendix.

3 Rank Inference in Fixed Effects Risk Adjustment Models

Risk adjustment is an important topic in performance evaluations of institutions, and discussions of risk adjustment models and related issues can be found in, for examples, Aitkin and Longford (1986), Goldstein and Spiegelhalter (1996), Iezzoni (2003) among others. In this section, we focus on to incorporate our developments of rank inference to fixed effects risk adjustment models, under which frequentist population ranks are well defined and intuitive. Comments on rank inference in random effects models are provided in Section 5.

Consider the following fixed effects risk adjustment model in a typical generalized linear model (GLM) setting,

$$h\{\mathrm{E}(y_{is})\} = \boldsymbol{\alpha}_i^T \boldsymbol{z}_{is} + \boldsymbol{\beta}^T \boldsymbol{x}_{is}, \qquad s = 1, 2, \ldots, n_i, \ i = 1, 2, \ldots, k \tag{15}$$

where h is a link function, parameters $\boldsymbol{\alpha}_i$ are specifically for the institution i and $\boldsymbol{z}_{is}$ are the corresponding regression covariates, parameters $\boldsymbol{\beta}$ are some additional but institution free regression coefficients and $\boldsymbol{x}_{is}$ are the corresponding regression covariates, including risk adjustment factors. The "model 2" of Aitkin and Longford (1986) is a Gaussian error fixed effects risk adjustment model. It is a special case of model (15) where h is the identity link function, $z_{is} \equiv 1$ and α_i is the specific intercept parameter (a scalar) for the institution i.

Let $\boldsymbol{z}^*$ be a vector of some pre-specified reference values for the covariates $\boldsymbol{z}_{is}$. For example, it can be the corresponding median or some representative values of $\boldsymbol{z}_{is}$'s across all institutions. Denote by $\zeta_i = \boldsymbol{\alpha}_i^T \boldsymbol{z}^*$, and they can be used to define population ranks. In particular, the conventional rank definition for the first institution is

$$R_1 = 1 + \sum_{j=2}^{k} G(\zeta_j - \zeta_1) = \sum_{j=2}^{k} \left\{ \mathbf{1}_{(\zeta_i < \zeta_1)} + \frac{1}{2} \mathbf{1}_{(\zeta_j = \zeta_1)} \right\}.$$

Let $\hat{\zeta}_i$ be a root-n consistent and asymptotically normally distributed estimator of ζ_i, for $i = 1, 2, \ldots, k$. Replacing ζ_j by $\hat{\zeta}_j$ in the rank R_1, we have a "plug-in" rank estimator

$$\hat{R}_1 = 1 + \sum_{j=2}^{k} G(\hat{\zeta}_j - \hat{\zeta}_1) = 1 + \sum_{j=2}^{k} \mathbf{1}_{(\hat{\zeta}_j \leq \hat{\zeta}_1)}.$$

In the special model "model 2" of Aitkin and Longford (1986), we have $\zeta_i = \alpha_i$. The R_1 and $\hat{R}_1$ defined above are exactly what Aitkin and Longford (1986) used for their "model 2".

As in Section 2, we can define a tie set or a near tie set to the first institution:

$$\Theta_T = \{j : \zeta_j = \zeta_1, \text{ for } j = 2, 3, \ldots, k\} \text{ or } \Theta_N = \{j : \sqrt{n}|\zeta_j - \zeta_1| = O(1), \text{ for } j = 2, 3, \ldots, k\}.$$

In the case of no ties with $\Theta_N = \Theta_T = \emptyset$, the result is trivial: $\hat{R}_1$ is a consistent estimator of R_1 and it could serve as a "singleton confidence set" for R_1 with asymptotically perfect coverage. As in Section 2, we are interested in the cases when there are potential ties or near ties. In these cases, the conventional rank estimator $\hat{R}_1$ is not consistent and behaves poorly. Under this context, we can again define a smooth population rank

$$R_1^{(smooth)} = 1 + \sum_{j=2}^{k} G_n(\zeta_j - \zeta_1) = 1 + \sum_{j=2}^{k} \bar{F}\left(\frac{\zeta_j - \zeta_1}{\tau_n}\right)$$

and its estimator

$$\hat{R}_1^{(smooth)} = 1 + \sum_{j=2}^{k} G_n(\hat{\zeta}_j - \hat{\zeta}_1) = 1 + \sum_{j=2}^{k} \bar{F}\left(\frac{\hat{\zeta}_j - \hat{\zeta}_1}{\tau_n}\right).$$

As in Section 2, $R_1^{(smooth)}$ and R_1 are asymptotically equivalent in absence of near ties ($\Theta_N - \Theta_T = \emptyset$); but they are different when there are near ties ($\Theta_N - \Theta_T \neq \emptyset$).

We can extend the results in Theorem 2.1-2.3 to the general risk adjustment model (15). Theorem 3.1 states that in the absence of near ties ($\Theta_N - \Theta_T = \emptyset$), the smooth sample rank $\hat{R}_1^{(smooth)}$ is a consistent estimator of the conventional rank R_1 and it can be used to produce confidence intervals for the conventional population rank R_1.

Theorem 3.1 *Assume model (15) holds and let* $\kappa_n \equiv \sqrt{n}\min_{j\notin\Theta_T, j>1}\Big|\zeta_j - \zeta_1\Big| \to \infty$. *Suppose the* $F(x)$ *and* τ_n *satisfy the conditions specified in Theorem 2.1.*
(i) Assume $\nu_1 = \{F'(0)\}^2 \mathrm{var}\big(\sum_{j\in\Theta_T}(\hat{\zeta}_j - \hat{\zeta}_1)\big) \neq 0$. *Then, as* $n \to \infty$, $\tau_n\sqrt{n}\,\nu_1^{-1/2}(\hat{R}_1^{(smooth)} - R_1)$ *is asymptotically distributed as* $N(0,1)$.
(ii) Suppose $\hat{\zeta}_j^*$ *are computed from a bootstrap sample such that* $n^{1/2}(\hat{\zeta}_1^* - \hat{\zeta}_1, \ldots, \hat{\zeta}_k^* - \hat{\zeta}_k)' \sim n^{1/2}(\hat{\zeta}_1 - \zeta_1, \ldots, \hat{\zeta}_k - \zeta_k)'$ *in limit, almost surely. Let* $\hat{R}_1^{(smooth)*} = 1 + \sum_{j=2}^k \bar{F}\big((\hat{\zeta}_j^* - \hat{\zeta}_1^*)/\tau_n\big)$. *Then, the bootstrap distribution of* $\tau_n\sqrt{n}(\hat{R}_1^{(smooth)*} - \hat{R}_1^{(smooth)})$ *consistently estimates the distribution of* $\tau_n\sqrt{n}(\hat{R}_1^{(smooth)} - R_1)$ *in the case when* $\nu_1 \neq 0$. *In the case when* $\nu_1 = 0$, *this bootstrap result still holds, provided that* $F = \Phi$ *and* $\min_{j\neq 1}|\zeta_j - \zeta_1|/(\tau_n^2 n^{1/2}) \to \infty$.

Theorem 3.2 below states that, in the presence of near ties with $\Theta_N - \Theta_T \neq \emptyset$, $\hat{R}_1^{(smooth)}$ is a consistent estimator of $R_1^{(smooth)}$ and it can be used to construct confidence intervals for $R_1^{(smooth)}$. In addition, these intervals can be utilized to obtain a set of conservative confidence intervals for the conventional rank R_1.

Theorem 3.2 *Assume model (15) holds and let* $\kappa_n^* \equiv \sqrt{n}\min_{j\notin\Theta_N, j>1}\Big|\zeta_j - \zeta_1\Big| \to \infty$. *Suppose the* $F(x)$ *and* τ_n *satisfy the conditions specified in Theorem 2.2.*
(i) Assume $\nu_{1n}^* = \{F'(0)\}^2 \mathrm{var}\big(\sum_{j\in\Theta_N}(\hat{\zeta}_j - \hat{\zeta}_1)\big) \neq 0$, *Then, as* $n \to \infty$, $\tau_n\sqrt{n}\,(\nu_{1n}^*)^{-1/2}(\hat{R}_1^{(smooth)} - R_1^{(smooth)})$ *is asymptotically distributed as* $N(0,1)$.
(ii) Let $\hat{\zeta}_j^*$ *and* $\hat{R}_1^{(smooth)*}$ *be the bootstrap estimates as in Theorem 3.1 (ii). Then, the bootstrap distribution of* $\tau_n\sqrt{n}(\hat{R}_1^{(smooth)*} - \hat{R}_1^{(smooth)})$ *consistently estimates the distribution of* $\tau_n\sqrt{n}(\hat{R}_1^{(smooth)} - R_1^{(smooth)})$ *in the case when* $\nu_{1n}^* \neq 0$. *In the case when* $\nu_{1n}^* = 0$, *this bootstrap result still holds, provided that* $F = \Phi$ *and* $\min_{j\neq 1}|\zeta_j - \zeta_1|/(\tau_n^2 n^{1/2}) \to \infty$.
(iii) The conventional rank estimator $\hat{R}_1$ *is not consistent to either* R_1 *or* $R_1^{(smooth)}$. *The*

difference between R_1 *and* $R_1^{(smooth)}$ *is*

$$R_1 - R_1^{(smooth)} = \sum_{j \in \Theta_N} \frac{1}{2} \{ \mathbf{1}_{(\zeta_j < \zeta_1)} - \mathbf{1}_{(\zeta_j > \zeta_1)} \} + o(1),$$

which is bounded between $-||\Theta_N||/2$ *and* $||\Theta_N||/2$. *However,* $\widehat{||\Theta_N||} = \sum_{j=2}^{k} K\big(\frac{\hat{\zeta}_j - \hat{\zeta}_1}{\tau_n}\big)$ *is a consistent estimator of* $||\Theta_N||$ *and a set of confidence interval for* R_1 *is*

$$floor\Big(L_1^* - \frac{1}{2}\widehat{||\Theta_N||}\Big) \leq R_1 \leq ceiling\Big(U_1^* + \frac{1}{2}\widehat{||\Theta_N||}\Big).$$

It is guaranteed that, asymptotically, the frequency coverage of this interval is more than $100\gamma\%$. *Here,* K *is the kernel function described in Section 2.3, and* L_1^* *and* U_1^* *are the lower and upper bounds of a* $100\gamma\%$ *confidence interval of the smooth rank* $R_1^{(smooth)}$*, respectively.*

The proofs of the theorems are similar to those of Theorem 2.1-2.3, noting that the standard root-n asymptotic normality applies to the estimators of the reference parameters ζ_i's in the GLM setting. The details are omitted.

4 Data Analysis Example

As part of the National Committee for Quality Assurance (NCQA) Quality Compass Report, also known as NCSA "report card", health facilities and organizations are evaluated on sets of composite measures of a variety of general prevention and disease-specific indicators (Schneider et al., 1999). Despite recommendations not to over-extrapolate the information in the publicly available report cards, health facilities, organizations and government agencies are "increasingly" ranked based on these types of report cards (Pogach et al., 2005). Pogach et al. (2005) used a cross-sectional analysis to illustrate the performance of the ranking method in identifying health care system's top and bottom performers, and found that the method had high uncertainty (sensitivity around 64%-83%). Part of the problem is that the conventional ranking approach does not go far enough to accommodate the uncertainty and random errors in the data. In this section, we use a data set used by Pogach et al. (2005) to illustrate our proposed ranking methods and confidence intervals for ranks.

Table 1 contains a data set used in Pogach et al's (2005) study. The first column of

Table 1 is the (fake) identification numbers of the 79 VHA facilities across the United States. The second column contains, as a performance measure, the percentage of diabetes patients in each facility who have poor Alc control ($\hat{p}_i$). The third column contains the total number of patients with diabetes in each facility (n_i). Although the database used by Pogach et al. (2005) also contained some demographical information such as age and gender, etc., it did not include any additional potential risk adjustment factors, such as severity of patients diabetes status, among others. To simplify our analysis and for the purpose of demonstrating our methodology, we will illustrate our method only using the data available in Table 1 and without including any risk adjustments. Based on the results in Section 3, we anticipate that the performance of the proposed ranking methods in the case of risk adjustment models behave similarly.

The ranks of the facilities are determined according to the parameters of their true poor A1c control rates p_i's or, equivalently, the log odds of poor Alc control rates $\theta_i = \log(p_i/(1-p_i))$'s. From Table 1 we can calculate the observed log odds $\hat{\theta}_{in} = \log(\hat{p}_i/(1-\hat{p}_i))$, which is asymptotically normally distributed with mean θ_i and variance $(1/p_i + 1/(1-p_i))/n_i$. Figure 1 is a scatter plot of the observed log odds $\hat{\theta}_i$ against the facility IDs (sorted by the observed $\hat{p}_i$). The short horizontal bars above and beneath each point indicate the upper and lower bounds of their point-wise 95% confidence intervals for the 79 facilities. Pairwise t-tests or z-tests of neighboring facilities have suggested that none of the log odds of the 79 facilities can be significantly separated from those of their neighboring facilities.

Four types of confidence intervals for population ranks are considered here and in Section 5. They are intervals obtained using 1) the conventional rank based centered bootstrap percentile method, 2) the conventional rank based regular percentile bootstrap method, 3) the smooth rank based bootstrap method as described in Sections 2.1 and 2.2, and 4) the smooth rank interval based adjusted method as described in Section 2.3. The conventional bootstrap methods 1) and 2) for general cases are described in Efron and Tibshirani (1994), where the centered bootstrap percentile method 1) yields a level γ confidence interval of formula $(2\hat{R}_i - b_{(1+\gamma)/2}, 2\hat{R}_i - b_{(1-\gamma)/2})$ and the basic bootstrap percentile method 2) yields an interval of formula $(b_{(1-\gamma)/2}, b_{(1+\gamma)/2})$. Here, b_t is the t-th quantile of the bootstrap distribution of the bootstrap estimator $\hat{R}_i^*$. To implement the smooth rank based methods 3) and 4) developed in the paper, we use $F(x) = \Phi(x)$ and a tuning parameter $\tau_n \propto n^{-\beta/2}$, for some $0 < \beta < 1$; details are described in the next two paragraphs.

Let s_{in}^2 be the sample estimate of $\mathrm{var}(\hat{\theta}_{in})$. Since n is the generic sample size representing the order of s_{in}^{-2} as described in Section 1, for the interval estimation of R_i we use $\tau_n = a\,\mathrm{s}_{in}^{\beta}$, where the constant $a = 0.29$ is the interquartile range of the 79 values of $\hat{\theta}_{in}$ to reflect the overall spread of θ_i's. When β varies from 0 to 1, this τ_n formula yields possible τ_n values typically ranging from 0.01 to 0.30, which is a wide enough tuning range for most facilities. The key task remained is to determine the constant β in the τ_n formula using the observed data. We employ the idea of choosing bandwidth in kernel estimation practices and consider a trade off to protect against over or under smoothing: (i) the proposed smooth ranks should not be too different from the conventional ranks and (ii) the smooth rank based confidence intervals obtained according to Theorems 2.1 and 2.2 should perform reasonably well with a close to $100\gamma\%$ coverage. In particular, we compare the trade-off of the MSEs of (i) the individually scaled rank differences $\mathrm{MSE}_1(\beta) = (1/79)\sum_{i=1}^{79}(R_i^{(smooth)} - R_i)^2/\{R_i(80-R_i)\}$ and (ii) miss-coverage $\mathrm{MSE}_2(\beta) = (1/79)\sum_{i=1}^{79}(c_i-\gamma)^2/\{\gamma(1-\gamma)\}$, where c_i is a frequency coverage percentage of the smooth rank confidence intervals from repeated experiments.

Assuming the true population A1c-out-of-control rates are as listed in Table 1, we plot in Figure 2 (a) and (b), $\mathrm{MSE}_1(\beta)$ in solid curves and $\mathrm{MSE}_2(\beta)$ in broken curves. Here, the coverage c_i's are computed from 1000 simulation experiments. We can see that smaller the constant β, better the performance of the smooth rank intervals, but bigger the differences between the smooth and conventional ranks. A min-max rule

$$\beta = \mathrm{argmin}_{0<\beta<1} \max\left\{\mathrm{MSE}_1(\beta),\ \mathrm{MSE}_2(\beta)\right\}$$

leads β to a range between 0.35 to 0.55, and for convenience we have picked $\beta = 0.5$ in our analysis; see Figure 2 (c)-(d). A similar conclusion can be reached by the minimax rule using $\mathrm{MSE}_3(\beta) = (1/79)\sum_{i=1}^{79}(R_i^{(smooth)} - R_i)^2$ and $\mathrm{MSE}_4(\beta) = (1/79)\sum_{i=1}^{79}(c_i-\gamma)^2$, provided that in the min-max comparison they are (externally) standardized by subtracting their respective medians and dividing by their respective robust scales. Figure 2 (e) and (f) further suggest a good trade-off between interval lengths and the MSE of under coverage

$\text{MSE}_5(\beta) = (1/79)\sum_{i=1}^{79}(c_i - \gamma)^2 \mathbf{1}_{(c_i<\gamma)}$ for the adjusted intervals proposed in Section 2.3.

Figure 3 (a)-(d) plot in solid points the conventional rank estimates against the 79 sorted facility IDs (sorted by the observed $\hat{p}_i$). These points are on the 45 degree line, as expected. The hollow circle points around the solid points in Figure 3 (c) correspond to the smooth rank estimates $\hat{R}_{in}^{(smooth)}$ of the 79 facilities. The vertical lines in Figure 3 (a)-(d) mark the 95% confidence intervals of the ranks for the 79 facilities. The confidence intervals are generated based on the aforementioned four methods 1)- 4), respectively. The confidence intervals in Figure 3 (b) are also asymptotic Bayesian credible intervals. The confidence intervals in Figure 3(c) are intended for the newly defined population rank $R_i^{(smooth)}$ and they can be used to cover the conventional rank R_i in the cases of no ties or true ties (but not the near ties). The confidence intervals in Figure 3 (d) are intended to cover the conventional ranks R_i in all cases. Roughly speaking, the intervals in Figure 3 (c) are the shortest, followed by the intervals in Figure 3 (a) and (b), which are followed by the intervals in Figure 3 (d). Without knowing the true parameter values θ_i's, we can not formally compare these rank intervals in terms of coverage probabilities. This will be carried out in simulation studies in the next section.

5 Simulation Studies

In order to examine the performance and operating characteristic of the proposed confidence intervals, we turn to simulation studies. In the first simulation study, we assume the $\hat{p}_i$ values listed in Table 1 are the "true" p_i values for the facilities, and calculate the "true" log odds θ_i values. Using the sample sizes listed in Table 1, we simulate an "observed" set of samples $\tilde{p}_i$ and compute the "observed" log odds $\tilde{\theta}_i$ values. In this simulation study, none of the "true" p_i or θ_i values are the same, but they are not distinguishable from some others at the current sample sizes by any tests, either. So, it is the case of near ties. Based on the simulated set of "observed" data, we construct the four types of confidence intervals for the "true" population ranks. The simulation is repeated 500 times. The coverage rate of these 500 confidence intervals from each of the four methods at levels 90% and 95% are plotted in Figure 4. For each facility and at each confidence level, we also obtain the median length of these corresponding 500 confidence intervals. The first panel of Table 2 contains the summary statistics of the 79 median lengths of the 79 facilities at level 90% and 95% respectively. Also, included in the the first panel of Table 2 are two summary statistics for the intervals achieving the target-coverage and accompanying under-coverage, which are defined by $\text{MSE}_4 = (1/79)\sum_{i=1}^{79}(c_i - \gamma)^2$ and $\text{MSE}_5 = (1/79)\sum_{i=1}^{79}(c_i - \gamma)^2\mathbf{1}_{(c_i<\gamma)}$ with the intended confidence level $\gamma = 90\%$ or 95%. A small MSE_4 suggests the coverage is close to the target level γ, while a small MSE_5 suggests there is no problem of under-coverage.

It is clear that the conventional rank based centered bootstrap percentile method is extremely poor in terms of coverage probabilities in this near ties case. The conventional rank based regular bootstrap percentile method behaves a little better, although it is still poor with some coverage as low as 67.6% and 76.0% at the intended levels 90% and 95% respectively. The smooth rank intervals have the right coverage for the smooth ranks with lengths slightly shorter than those of the conventional rank based methods. The conservative smooth rank adjusted intervals over-cover the conventional ranks, as expected. These results clearly match the results in Section 2, but not that in equation (3).

In the second simulation study, we repeat what we did in the first simulation study, after we artificially increase the sample sizes 20,000 times in each hospital. By drastically increasing the sample sizes, we are able to distinguish most of the "true" parameter values $\hat{p}_i$'s and $\hat{\theta}_i$'s, for example, using the regular t-test. So, this creates a no tie situation (almost, see discussions later). In this case of a well separated θ_i, $\hat{R}_i = R_i$ for a large n, as stated in (3). As in Figure 4, the coverage probabilities of the four types of intervals are plotted (figure is not included in the paper). From the figure, the smooth rank based confidence intervals again have the intended coverage but the rest three types of intervals have 100% coverage regardless of the confidence levels, except that there are still a few cases in the centered bootstrap percentile method with low coverage. The second panel of Table 2 contains the summary statistics of the median lengths and coverages (from 500 simulations) of the rank confidence intervals for the 79 facilities. Again, the results are as expected and match with our understanding in the no tie case.

In the third simulation, we create a true ties situation by grouping the 79 facilities into one group of 9 facilities and seven groups of 10 facilities. For each of the 9 or 10 facilities in a group, their true θ_i values are set to be the median value of the θ_i's in that group. To make the groups separated from one another, we again increase the sample sizes of each facility 20,000 times as in the second simulation. We carry out the simulation study parallel to the first or the second simulation study. As in Figure 4, the coverage probabilities of the rank intervals are plotted (figure not shown in this paper). From the figure, the conventional bootstrap method 1) has terrible coverage for the intended true population ranks. The conventional rank method 2) (equivalently Bayesian method) and the smooth rank based method 3) all have the right coverage. The smooth rank based adjusted intervals by 4) over cover. The third panel in Table 3 contains the summary statistics of the median lengths and frequency coverages (from 500 simulations) of the 79 confidence intervals for ranks in this true ties case. Clearly, the intervals created from the conventional bootstrap methods (including the asymptotic Bayesian method) are very long. They are even longer than those conservatively adjusted intervals by the method 4), which cover the true conventional rank almost all the time.

In the second and third simulation studies, we increase the sample size 20,000 times, resulting more than 20 millions samples in each facility! These unusually huge sample sizes are unrealistic. We include these two simulation studies only to illustrate the concept and theory. In fact, we first tried to the increase of the sample size 5,000 and then 10,000 times in the last two simulations, and didn't get intended results suggested by the theories for the no tie and true tie cases. By examining the data, we realize that many of the rates listed in Table 1 are extremely close. For example, facility 31 and 32 have the $\hat{p}_i$ rates .1010 and .1011 and sample sizes 2940 and 3649 respectively. Assuming these are their true p_i rates, it requires a sample size of 20,000 times to barely tell them apart by the regular z or t-test. This observation also reenforces our belief that the tie and near tie assumption may be appropriate for the data in Table 1 and also many other data in practice. In this case, the newly proposed smooth rank approach is preferred. One nice thing about the newly proposed approach is that they work well in all situations, including the no ties, true ties, and near ties cases.

6 Discussions

In this paper, we developed rigorous frequentist inference theory based on nonstandard bootstrap for population ranks, especially for the cases with ties and near ties. The conventional rank parameter is known to be a "difficult" parameter in the literature, in terms of making inference. When there are ties or near ties, the conventional sample rank is not even consistent and inference based on it behaves very poorly. We proposed a smooth sample rank which can be used to derive inference on the conventional rank parameter under the assumption of no ties or only true ties. In the case of near ties, we provide a smooth population rank definition which has built-in sample sizes. This smooth population rank can be consistently estimated by the aforementioned smooth sample rank, and the results on the smooth ranks also can be utilized to make inference for the conventional rank. The results are extended to a general fixed effects risk adjustment model.

In the presence of near ties, we have also tried to explore several alternatives to obtain confidence bounds for the conventional rank. One alternative, utilizing the fact that R_1 is discrete and integer-valued, is simple, theoretically sound for large sample, but failed to provide desirable coverage for the example of Table 1 where there are many ties or near ties. The method is as follows. Let $S(t)$ be a nondecreasing uniformly continuously differentiable distribution function with $S(t) = 1$ for $t \geq 0$ and let τ_n be as in (9). Define $\tilde{U}_1 = 1+\text{round}\left(\sum_{j=2}^{k} S\left(\frac{\hat{\theta}_{1n}-\hat{\theta}_{jn}}{\tau_n}\right)\right)$ and $\tilde{L}_1 = k-\text{round}\left(\sum_{j=2}^{k} S\left(\frac{\hat{\theta}_{jn}-\hat{\theta}_{1n}}{\tau_n}\right)\right)$, where $\text{round}(x)$ is the integer nearest to x. It is not hard to prove (proof omitted) that, as $n \to \infty$, $P\left\{\tilde{L}_1 \leq R_1 \leq \tilde{U}_1\right\} \to 1$. Unfortunately, these bounds are too tight and could not provide decent coverage for the example of Table 1. A close examination of the proof reveals that, in the case of many ties or near ties, it requires to have huge sample sizes n_i so that an $o_p(1)$ term could be bounded by $\frac{1}{2}$. This observation is confirmed by a simulation with artificially increased sample sizes n_i. We also have considered another conservative approach utilizing simultaneous intervals. In that approach, we obtain a re-sampling based simultaneous confidence region for the parameter vector $(\theta_1, \theta_2, \ldots, \theta_k)^T$, then project it on the rank space to obtain confidence intervals for the conventional ranks. Although the simultaneous coverage is guaranteed, this procedure produces too wide confidence intervals for single ranks. For practical examples with a large number of ties or near ties, we still recommend the use of the adjusted smooth rank intervals proposed in Section 2.3 to obtain confidence bounds for the conventional rank.

Random effects models are also common in risk adjustment practice. To compare with their fixed effects "model 2", Aitkin and Longfold (1996) used a random effects "model 5"

$$y_{is} = b_i + \boldsymbol{\beta}^T \boldsymbol{x}_{is} + \epsilon_{is}, \qquad s = 1, 2, \ldots, n_i,\ i = 1, 2, \ldots, k, \tag{16}$$

where random intercept $b_i \sim N(\alpha, \sigma_b^2)$ and random error $\epsilon_{is} \sim N(0, \sigma^2)$. Under (16), they define the rank of the first institution as $\tilde{R}_1 = 1 + \sum_{j=2}^{k} \mathbf{1}_{(b_i < b_1)}$, which is a random quantity. Bayesian type of analysis for $\tilde{R}_i$ seems appropriate, but it is not a well defined parameter from a frequentist viewpoint. Note that b_i are random draws from $N(\alpha, \sigma_b^2)$ and a replicate study likely has a different set of k institutions. The frequentist developments in this paper, treating rank as a fixed parameter, can not directly be applied to such $\tilde{R}_1$.

The developments in this paper can be extended to other random effects models, however. Suppose the k (assume now a large number) institutions are sampled from a finite number (say m) of sub-population or groups. For example, the sub-populations can be schools or hospitals in different regions or of different types. We are interested in ranking these m sub-populations. The random effects model (16), in this case, is

$$y_{ils} = b_{il} + \boldsymbol{\beta}^T \boldsymbol{x}_{ils} + \epsilon_{ijs} \qquad s = 1, 2, \ldots, n_{il},\ l = 1, 2, \ldots, k_i, i = 1, 2, \ldots, m.$$

with random intercept $b_{il} \sim N(\alpha_i, \sigma_b^2)$, normal error $\epsilon_{ils} \sim N(0, \sigma^2)$ and $k = \sum_{i=1}^{m} k_i$. In this case, the population ranks for the m sub-populations are well defined $R_i = 1 + \sum_{j \neq i} \left\{ \mathbf{1}_{(\alpha_j < \alpha_i)} + \frac{1}{2} \mathbf{1}_{(\alpha_j = \alpha_i)} \right\}$ and $R_i^{(smooth)} = 1 + \sum_{j=2}^{k} G_n(\alpha_j - \alpha_i)$, $i = 1, 2, \ldots, m$. The developments of this paper can be directly applied to such random effects models.

APPENDIX

Proof of Theorem 2.1. It follows from (10) that

$$\max_{j\notin\Theta_T, j\geq 2}\Big|F\big((\theta_j-\theta_1)/\tau_n\big)-\mathbf{1}_{(\theta_j>\theta_1)}\Big| = o\Big(n^{-1/2}/\tau_n\Big) \tag{A.1}$$

and $F'\big((\theta_j-\theta_1)/\tau_n\big) \to 0$ for $j \notin \Theta_T$, since the uniform differentiability of F implies $F'(x)\to 0$ as $F(x)\to 0$ or $1-F(x)\to 0$. By (5) and (9), $\frac{\hat\theta_{jn}-\theta_j}{\tau_n} = \frac{Z_{jn}+o_P(1)}{\tau_n\sqrt{n}} = o_P(1)$. Thus, Taylor expansion yields

$$\begin{aligned}
&\hat R_1^{(smooth)} - \Big\{k - \sum_{j=2}^{k} F\Big((\theta_j-\theta_1)/\tau_n\Big)\Big\} \qquad\qquad (A.2)\\
=& -\sum_{j=2}^{k}\Big\{F'\Big((\theta_j-\theta_1)/\tau_n\Big)+o_P(1)\Big\}\Big(\frac{\hat\theta_{jn}-\theta_j}{\tau_n}-\frac{\hat\theta_{1n}-\theta_1}{\tau_n}\Big)\\
=& -\sum_{j\in\Theta_T}\frac{F'(0)}{\tau_n n^{1/2}}\Big(Z_{jn}-Z_{1n}+o_P(1)\Big) \sim \frac{N(0,v_1)+o_P(1)}{\tau_n n^{1/2}}.
\end{aligned}$$

The asymptotic normality in Part (i) follows since (1) and (A.1) imply

$$R_1 - \Big\{k-\sum_{j=2}^{k}F\Big((\theta_j-\theta_1)/\tau_n\Big)\Big\} = \sum_{j\notin\Theta_T}\Big\{F\Big((\theta_j-\theta_1)/\tau_n\Big)-\mathbf{1}_{(\theta_j>\theta_1)}\Big\} = o_P(n^{-1/2}/\tau_n).$$

Part (ii) of the theorem is derived from a similar expansion of $\hat R_1^{(smooth)*} - \hat R_1^{(smooth)}$ when $v_1\neq 0$. When $v_1 = 0$, further expansions of $\hat R_1^{(smooth)} - \hat R_1$ and $\hat R_1^{(smooth)*} - \hat R_1^{(smooth)}$ are needed under $F=\Phi$; details are omitted. $\diamond$

Proof of Theorem 2.3. Note that $K(0) = 1$, $K(x)\to 0$ as $|x|\to\infty$, and $(\hat\theta_j - \hat\theta_1)/\tau_n \to 0$ if $j\in\Theta_N$ and $(\hat\theta_j-\hat\theta_1)/\tau_n\to 1$ if $j\notin\Theta_N$. We have

$$\widehat{||\Theta_N||} = \sum_{j=2}^{k} K\Big(\frac{\hat\theta_j-\hat\theta_1}{\tau_n}\Big) = \sum_{j\in\Theta_N} K\Big(\frac{\hat\theta_j-\hat\theta_1}{\tau_n}\Big) + o_p(1) = ||\Theta_N|| + o_p(1).$$

The first result of the theorem follows immediately.

Since L_1^* and U_1^* are the lower and upper confidence bounds of the smooth rank $R_1^{(smooth)}$, it is easy to see from (14) that

$$P\Big(L_1^* - \frac{1}{2}\widehat{||\Theta_N||} + o_p(1) \leq R_1 \leq U_1^* + \frac{1}{2}\widehat{||\Theta_N||} + o_p(1)\Big) \geq 1-\beta.$$

Note that R_1 is a discrete integer number. Also, when $n \to \infty$, floor$(L_1^* - \frac{1}{2}\widehat{||\Theta_N||}) \leq L_1^* - \frac{1}{2}\widehat{||\Theta_N||} + o_p(1)$ and ceiling$(U_1^* + \frac{1}{2}\widehat{||\Theta_N||}) \geq U_1^* + \frac{1}{2}\widehat{||\Theta_N||} + o_p(1)$, in probability. The second result of the theory follows immediately. $\diamond$

Reference

Aitkin, M. and Longford, N. (1986). Statistical modelling issues in school effectiveness studies (with discussion). *J.R. Statist. Soc.* A. 149, 1-42.

Austin, P.C. and Tu, J.V. (2006). Comapring clinical data with administrative data for producing acute myocardinal report cards. *J.R. Statist. Soc.* A. 169, 115-126.

Christiansen, C.L. and Morris, C.N. (1997). Improving the statistical approach to health care provider profiling. *Annals of Internal medicine.* 127, 764-768.

Efron B. and Tibshirani R.J. (1994). *An Introduction to the Bootstrap.* Chapman & Hall.

Goldstein, H. and Spiegelhalter, D.J. (1996). League tables and their limitations: statistical issues in comparisons of institutional performance (with discussion). *J.R. Statist. Soc.* A. 159, 385-443.

Iezzoni, L.I. (2003) (ed.). *Risk Adjustment for Measuring Health Care Outcomes*, 3rd edition. Health Administration Press.

Laird, N.M. and Louis, T.A. (1989). Empirical Bayes ranking models. *J. of Educational Statistics.* 14, 29-46.

Landrum, M.B., Normand, S-L. T., and Rosenheck, R.A. (2003). Selection of related multivariate means: monittoring psychiatric care in the department of vaterans affairs. *J. Am. Statist. Assoc..* 98, 7-16.

Lockwood, J.R., Louis, T.A., McCaffrey, D.F. (2002). Uncerteinty in rank estimation: implications for value-added modeling accountability systems. *J. of Educational and Behavioral Statistics.* 27, 255-270.

Lin, R., Louis, T.A., Paddock, S.M. and Ridgeway, G. (2006). Loss function based ranking in two stage hierarchical models. *Bayesian Analysis.* 1, 915-946.

Marchand, E. and Strawderman, W.E. (2006). On the behavior of Bayesian credible intervals for some restricted parameter space prblems. IMS Lecture Notes.

McCaffrey, D.F., Lockwood, JR., Koretz, D., Louis, T.A., and Hamilton, L. (2004). Models for value-added modeling of teacher effects. *J. of Educational and Behavioral Statistics.* 29, 67-101.

Noell, G.H. and Burns, J.L. (2006). Value-added assessment of teacher preparation: an illustration of emerging technology. *Journal of Teacher Education.* 57, 37-50.

Normand, S-L., Glickman, M.E. and Gatsonis, C.A. (1997). Statistical methods for profiling providers of medical care: issues and applications.*J. Am. Statist. Assoc.*92,803-814.

Pogach L, Xie M, Shentue Y, Tseng CL, Maney M, Rajan M, Tiwari A, Kolassa J, Helmer D, Crystal S, Safford M. (2005). Diabetes Healthcare Quality Report Cards: How Accurate Are The Grades? *American Journal of Managed Care.* 11, 797-804.

Rubin, D.B., Stuart, E.A., and Zanutto, E.L. (2004). A potetial outcomes view of value-added assessment in education. *J. of Educational and Behavioral Statistics.* 29, 103-116.

Schneider EC, Riehl V, Courte-Wienecke S, Eddy DM, Sennett C. (1999). Enhancing performance measurement: NCQA's road map for a health information framework. *J. Am. Med. Assoc.*. 282, 1184-1190.

Shen, W. and Louis, T.A. (1998). Triple-goal estimates in two-stage hierachical models. *J.R. Statist. Soc.* B. 60, 455-471.

Snijders, T.A.B. (1996). Discussion on "League tables and their limitations: statistical issues in comparisons of institutional performance" by Goldstein and Spiegelhalter. *J.R. Statist. Soc.* A. 159, 435.

Thompson, W., Wang, H., Xie, M., Kolassa J., Rajan M., Tseng C.L., Crystal S., Zhang Q., Vardi Y., Pogach L., Safford M.M. (2005). Assessing Quality Of Diabetes Care By Measuring Longitudinal Changes In Hemoglobin A1c in The Veterans Health Administration. *Health Services Research Journal.* 40, 1818-1835.

Zhang, T. and Woodroofe, M. (2002). Credible and confidence sets for restricted parameter spaces. *J. Statist. Plann. Inference*, 115, 479-490.

Table 1. "Poor HbAlc Control" data of 79 VA facilities

Fac. ID	Poor Alc contr. rate $\hat{p}_i$	No. of cases n_i	Fac. ID	Poor Alc contr. rate $\hat{p}_i$	No. of cases n_i	Fac. ID	Poor Alc contr. rate $\hat{p}_i$	No. of cases n_i	Fac. ID	Poor Alc contr. rate $\hat{p}_i$	No. of cases n_i
1	.0444	1803	21	.0939	3811	41	.1057	2128	61	.1247	13598
2	.0612	2270	22	.0940	6641	42	.1063	1176	62	.1251	1758
3	.0673	1188	23	.0943	5577	43	.1065	6828	63	.1255	3020
4	.0700	4143	24	.0946	2252	44	.1066	2851	64	.1256	3624
5	.0706	2195	25	.0953	6902	45	.1068	3810	65	.1260	3254
6	.0733	2443	26	.0956	1015	46	.1072	2417	66	.1276	2719
7	.0738	2290	27	.0969	1558	47	.1078	2365	67	.1291	2779
8	.0799	1827	28	.0978	2208	48	.1099	1519	68	.1292	1648
9	.0803	4842	29	.0982	1293	49	.1100	2081	69	.1301	4573
10	.0839	3170	30	.0998	1423	50	.1106	4468	70	.1313	3230
11	.0846	2234	31	.1010	2940	51	.1122	4037	71	.1456	3380
12	.0848	1780	32	.1011	3649	52	.1123	3685	72	.1472	3044
13	.0851	2069	33	.1017	3620	53	.1132	3808	73	.1474	2952
14	.0854	3032	34	.1029	1078	54	.1134	2751	74	.1477	3576
15	.0863	3072	35	.1030	2990	55	.1137	5436	75	.1502	3182
16	.0897	2108	36	.1035	1266	56	.1147	1482	76	.1521	1407
17	.0927	7853	37	.1040	4502	57	.1161	4668	77	.1666	3812
18	.0929	2626	38	.1045	4513	58	.1169	1950	78	.1680	2321
19	.0933	1640	39	.1046	3059	59	.1203	4672	79	.1718	1624
20	.0935	1733	40	.1055	2284	60	.1231	5785			

Table2. Summary Statistics of the median lengths and coverages of the rank confidence intervals for 79 facilities in the simulations.

Panel 1: Near Ties Case (sample sizes = n_i)									
		1) Conv. Centered		2) Conv. Perc.		3) Smooth Rank		4) Smooth Rank Adj.	
		90%	95%	90%	95%	90%	95%	90%	95%
Length	Min.	0.00	0.00	0.00	0.00	0.29	0.51	2.00	2.00
	1st Qu.	12.50	14.50	12.50	14.50	11.99	14.33	16.50	18.50
	Median	18.00	22.00	18.00	22.00	17.87	21.47	23.00	26.00
	Mean	18.57	21.96	18.57	21.96	17.83	21.10	23.30	26.61
	3rd Qu.	25.00	29.50	25.00	29.50	23.74	28.03	31.00	35.50
	Max.	38.00	43.00	38.00	43.00	35.43	41.04	44.00	50.00
Coverage	MSE_4	0.9601	0.9904	0.1690	0.0840	0.0447	0.0244	0.3902	0.1031
Coverage	MSE_5	0.9340	0.9841	0.1140	0.0644	0.0410	0.0220	0.0061	0.0016

Panel 2: No Ties Case (sample sizes = 20,000n_i)									
		1) Conv. Centered		2) Conv Perc.		3) Smooth Rank		4) Smooth Rank Adj.	
		90%	95%	90%	95%	90%	95%	90%	95%
Length	Min.	0.0000	0.0000	0.0000	0.0000	0.00000	0.0000	0.0000	0.0000
	1st Qu.	0.0000	0.0000	0.0000	0.0000	0.0558	0.0667	1.0000	1.0000
	Median	0.0000	0.0000	0.0000	0.0000	0.1847	0.2201	2.000	2.0000
	Mean	0.1519	0.1772	0.1519	0.1772	0.1751	0.2087	1.5820	1.6200
	3rd Qu.	0.0000	0.0000	0.0000	0.0000	0.2769	0.3302	2.0000	2.0000
	Max.	1.0000	1.0000	1.0000	1.0000	0.5665	0.6748	3.0000	3.0000
Coverage	MSE_4	0.9795	0.6752	0.7687	0.1926	0.0367	0.0132	0.7900	0.1975
	MSE_5	0.3246	0.5143	0.0000	0.0000	0.0092	0.0038	0.0000	0.0000

Panel 3: True Ties Case (sample sizes = 20,000n_i)									
		1) Conv. Centered		2) Conv. Perc.		3) Smooth Rank		4) Smooth Rank Adj.	
		90%	95%	90%	95%	90%	95%	90%	95%
Length	Min.	5.0000	6.0000	5.0000	6.0000	0.7501	0.8931	6.0000	6.0000
	1st Qu.	6.0000	7.0000	6.0000	7.0000	0.9084	1.0810	6.0000	6.0000
	Median	7.0000	8.0000	7.0000	8.0000	0.9930	1.1810	6.0000	6.0000
	Mean	6.8730	7.6200	6.8730	7.6200	0.9961	1.1850	6.0000	6.3860
	3rd Qu.	7.0000	8.0000	7.0000	8.0000	1.0610	1.2620	6.0000	7.0000
	Max.	8.0000	9.0000	8.0000	9.0000	1.2600	1.4980	6.0000	7.0000
Coverage	MSE_4	19.4443	23.2261	0.0537	0.0195	0.0182	0.0077	0.7900	0.1975
	MSE_5	19.4443	23.2261	0.0021	0.0008	0.0103	0.0034	0.0000	0.0000

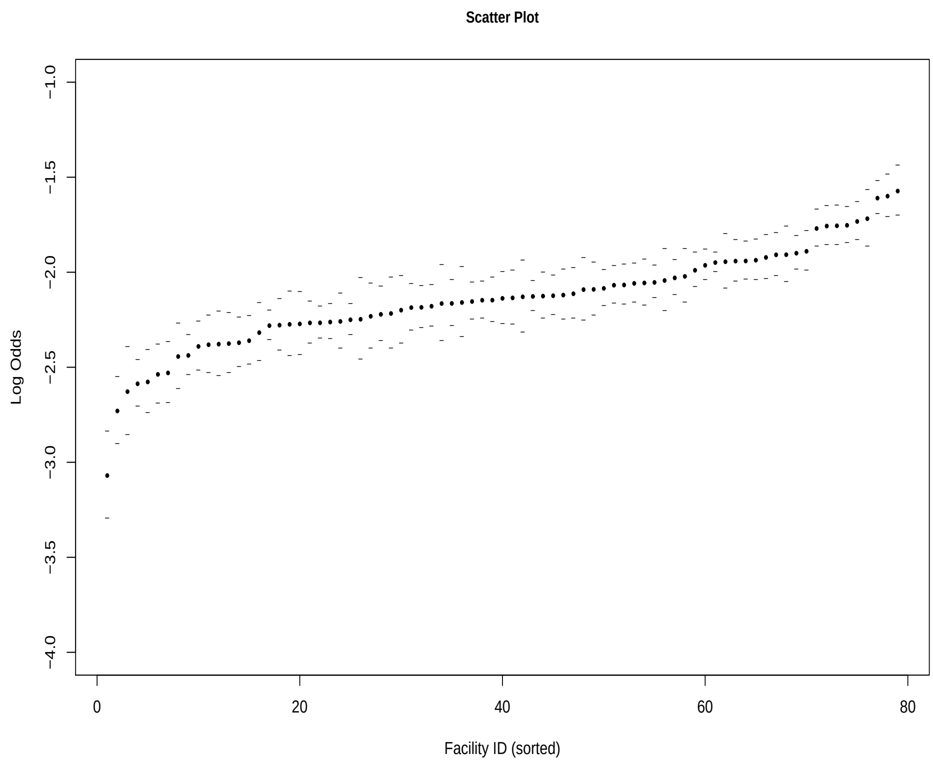

Figure 1: Figure 1 is a scatter plot of the observed log odds ratio $\hat{\theta}_{in}$ against the facility IDs (sorted by the observed $\hat{p}_i$). The short horizontal bars in the plot indicate the upper and lower bounds of the point-wise 95% confidence intervals for the 79 facilities. For any facility, there exist at least one other facility such that their normal theory based 95% confidence intervals overlaps. The 79 log odds are very close and cannot be distinguished from each other.

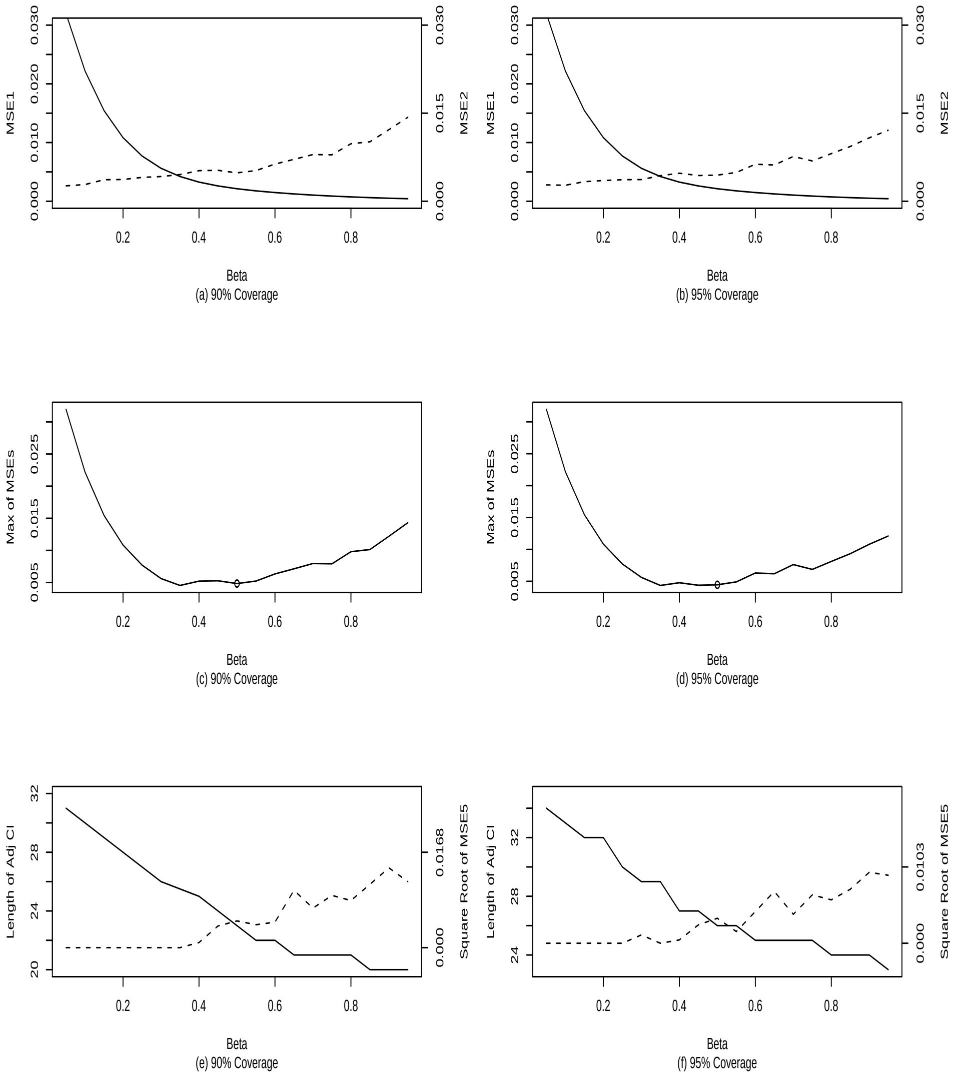

Figure 2: This figure illustrates the choice of β by the min-max rule described in Section 4. Figure 2 (a)-(b) show, at confidence levels $\gamma = 90\%$ and 95%, the opposite trends of $\text{MSE}_1(\beta)$ (in solid curves with scales marked on left side) and $\text{MSE}_2(\beta)$ (in dashed curves with scales marked on the right hand side). Figure 2 (c)-(d) plot, at confidence levels $\gamma = 90\%$ and 95%, the maximum of the two MSEs $\max\{\text{MSE}_1(\beta), \text{MSE}_2(\beta)\}$ at a β value, against β. The hollow circle points correspond to the choice of $\beta = 0.5$. Figure 2 (e) and (f) show, at confidence levels $\gamma = 90\%$ and 95%, the opposite trends of median interval lengths (in solid curves with scales marked on left side) and $\{\text{MSE}_5(\beta)\}^{.5}$ (in dashed curves with scales marked on the right hand side) as β increases from 0 to 1.

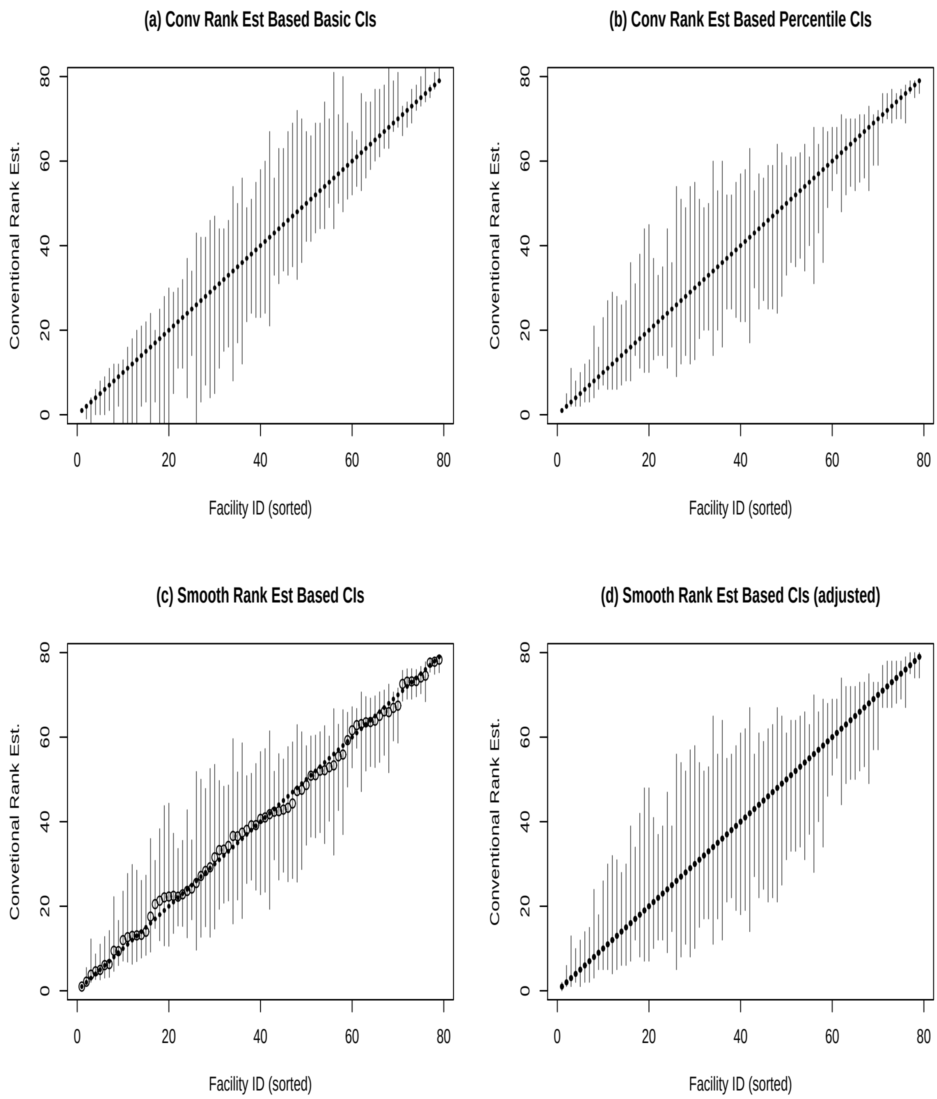

Figure 3: Figure 3 (a)-(d) plot in solid points the conventional rank estimates against the facility id's (sorted by the observed $\hat{p}_i$). The hollow circle points in Figure 3 (c) are the smooth rank estimates $\hat{R}_{in}^{(smooth)}$'s against the facility ids. The vertical lines mark the 95% confidence intervals of the ranks for the 79 facilities. The intervals in Figure 3 (a) are based on the conventional rank estimator using the centered bootstrap percentile method. The intervals in Figure 3 (b) are based on the conventional rank estimator using the regular bootstrap percentile method. Equivalently, they also correspond to an asymptotic Bayesian method. The intervals in Figure 3 (c) are based on the newly proposed smooth rank estimator $\hat{R}_{in}^{(smooth)}$, as described in Section 2.2. The intervals in Figure 3 (d) are the conservatively adjusted intervals described in Section 2.3 for the situation of near ties.

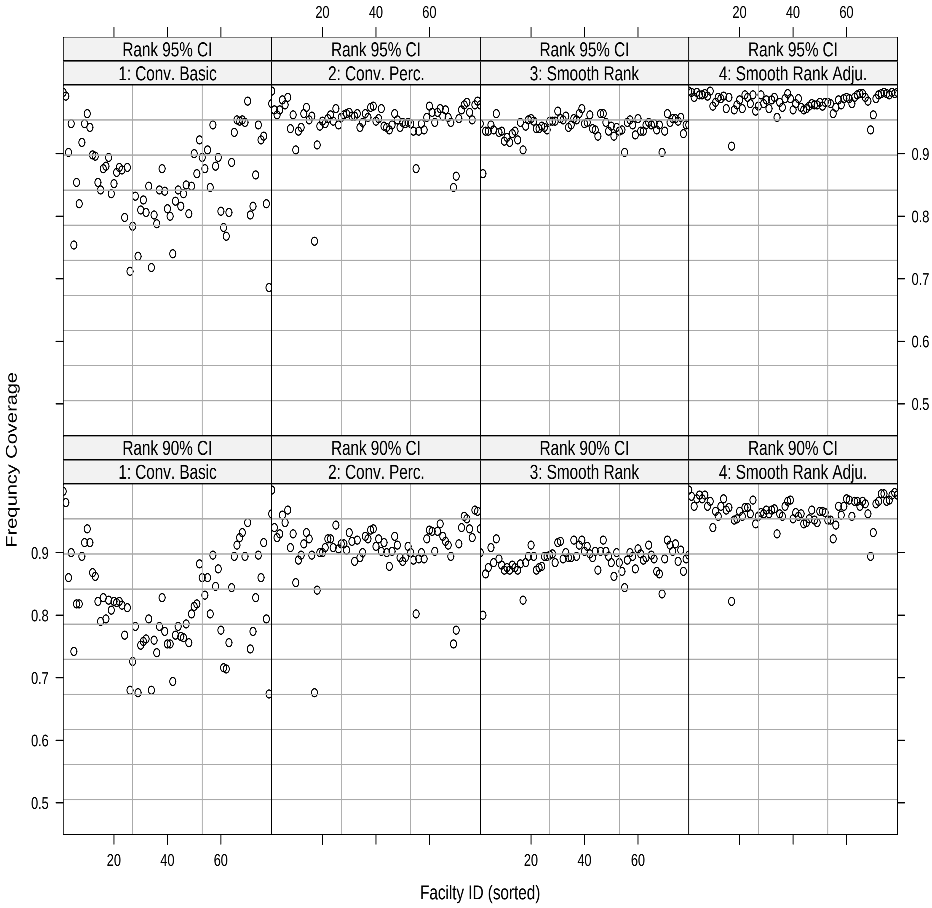

Figure 4: This figure is a Trellis plot used to examine the frequency coverage of four types of 90% and 95% rank confidence intervals in the case of existing near ties ($\Theta_N \neq \emptyset$). The four types of intervals are obtained from 1) the conventional rank based centered bootstrap percentile method, 2) the conventional rank based regular bootstrap percentile method, 3) the smooth rank based bootstrap method (for the smooth population rank $R_i^{(smooth)}$), and 4) the adjusted smooth rank based intervals (adjusted for the conventional population rank R_i). The first two methods have poor coverage and the fourth method has higher coverage than intended in the vast majority of cases.